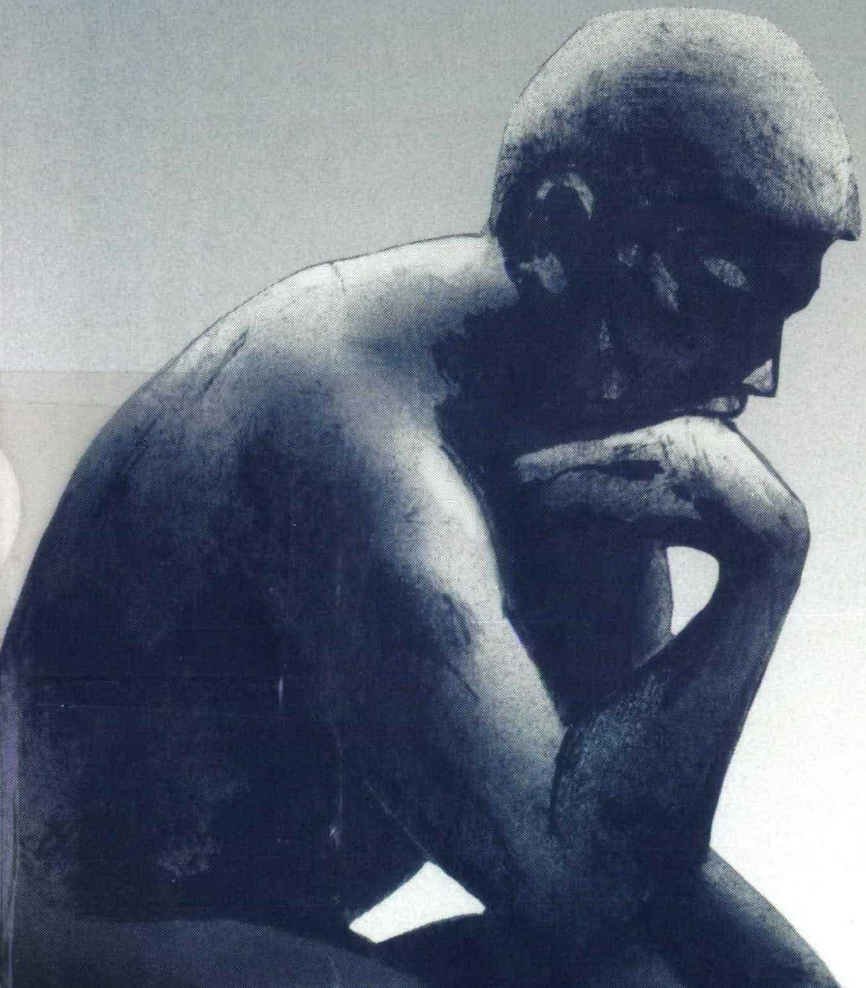
Partial logic and knowledge representation

Thijsse, Elias George Coenraad

Elias Thijsse

PARTIAL LOGIC AND KNOWLEDGE REPRESENTATION



Partial Logic and Knowledge Representation

Partial Logic and Knowledge Representation

Proefschrift ter verkrijging van de graad van doctor aan de Katholieke Universiteit Brabant, op gezag van de rector magnificus, prof. dr. L.F.W. de Klerk, in het openbaar te verdedigen ten overstaan van een door het college van dekanen aangewezen commissie in de aula van de Universiteit op woensdag 1 juli 1992 te 16.15 uur.

door

Elias George Coenraad Thijsse

geboren te Delft

promotoren:

prof. dr. J.F.A.K. van Benthem
prof. dr. H.C. Bunt

Eburon Publishers, P.O. Box 2867, 2601 CW Delft;
phone +3115 131484, fax +3115 146888

ISBN 90-5166-267-X

cover design by ir. M. Verweij

Acknowledgements

It is only fair to say that the work on this thesis would probably not have started without Harry Bunt and would not have been completed without Johan van Benthem. I would like to thank Harry for making the ITK Research Institute largely what it is, a stimulating environment with sufficient expertise in logic, language and knowledge representation to support the kind of research reported here, and for the amount of freedom I have experienced while working on this project.

After a stage in which the project focussed on semantico-pragmatic aspects of knowledge (chapter 5 is still reminiscent of this), there was a reorientation of the enterprise towards human knowledge in general (cf. chapter 6) and one way to deal with knowledge in computer systems (chapter 8). Then it turned out that *partial logic* might be a better candidate, both for describing the limited inferential power of human beings, and for reducing the size of the formal models of knowledge as needed for computer implementation. In order to increase my own understanding of this logic, and to make it fit the subject of knowledge representation, partial logic had to be developed further. Here in particular, Johan van Benthem enters the picture, although he was never really out of it. His guidance, speed and positive thinking, provided enough stimulus to carry on, eventually leading to an acceleration in the last few years. Needless to say this book benefited tremendously by his insight and knowledge of (modal) logic.

Though writing a dissertation is sometimes a solitary activity, the regular contacts with colleagues in Amsterdam, Tilburg and other cities meant a lot to me. I gratefully acknowledge my debt to René Ahn [chapter 6], Kees van Deemter [5], Jan van Eijck [5], Joseph Halpern [6,8], Jan Jaspars [2,6], Dick de Jongh [8], Aad Lander [1–5], John-Jules Meyer [1–6], Reinhard Muskens [2,4,6], Maarten de Rijke [2], Edith Spaan [9], Martin Stokhof [5], Frank Veltman [4] and Heinrich Wansing [6]. for their valuable comments and help.

I am also grateful for the invitations from abroad, which led me to the TARK1 and TARK3 conferences in Monterey (California), the SRI International Research Lab. in Cambridge (United Kingdom), the Gruppe für Logik, Wissenstheorie und Information an der Freie Universität Berlin, and the Nordic Graduate Studies Course in Linköping (Sweden).

Apart from the contents, a number of people contributed in other ways to the completion of this thesis. Within the period in which this research took place, Jan van Eijck let me join his work on a textbook on logic for language and information science. Besides instructing me how to use LaTeX, this also taught me how to organize book-length documents and meet deadlines in good temper. I am indebted to Henk Kempff for replacing me at a number of occasions when time was pressing. Martin Kammler and Anita Cremers assisted in preparing the paper underlying chapter 8; Petra Tuunter provided a number of pictures and instructed me how to IDRAW myself. Pieter Nieuwint corrected the English of papers underlying several chapters of this thesis; his corrections went well beyond the call of duty. Jan Jaspars, Tijn Borghuis, Emiel Krahmer and René Ahn jointly also did a great job proof-reading the manuscript. Since the manuscript went through a number of revisions, none of the people acknowledged should be held responsible for any remaining errors.

Apart from those who helped in some way or other to make the thesis what it is, I am grateful to my family, friends and colleagues for taking a keen interest in the enterprise. I would like to thank, in particular, my wife Jet for her perseverance, and my daughters Ilse and Sylvie for their joyful presence.

It almost goes without saying that what I am is to some extent the product of education. Going back in time, I would like to mention for that reason the lecturers of the departments of Philosophy, General Linguistics and Mathematics of the University of Groningen, and the stimulating atmosphere there during what has been called the 'semantic wave' in the early eighties. Before that, there were the teachers of the Christelijk Lyceum Delft who gave me the feeling I was perhaps not entirely hopeless.

But my personal history begins, of course, with the opportunities I was given by my late parents and, later on, the support offered by my parents-in-law. I therefore dedicate this book to those I cannot thank personally anymore: my parents Coen Thijsse and Iet Thijsse–Jongeneel, and my late father-in-law, Max van Everdingen.

Contents

Chapter 0

General introduction and overview

As the title of this thesis suggests we focus on the interrelated notions *partiality* and *knowledge*. Both the representation of human knowledge and the representation of knowledge in computers systems is expected to benefit from partial logic. The aim of this thesis is to provide a *systematic* study of partial logic, which will be developed further in order to be applicable to knowledge. We will first introduce the notion of 'partiality' below, motivate its incorporation into logic and argue for its intimate connection to knowledge. From this overall relation between partial logic and knowledge representation some major research questions are obtained. This introduction closes with an overview of the thesis: its themes, its parts, its chapters.

Partiality

In classical logic propositions are either true or false. In other words, *truth* and *falsity* are the classical truth values. Partial logic deviates from this: the truth value can be left open, i.e. the proposition can be undefined. There are quite a lot reasons for introducing *partiality* into logic:[1]

- lack of information: we simply may have insufficient knowledge to decide whether or not a proposition is true; consequently, the truth of the sentence

  (1) Mary works.

  may be undetermined, from an agent's point of view.

- lack of assertoric value: linguistic examples like (1) are mentioned rather than used in their normal meaning.

- solving semantic paradoxes: a sentence such as (2) cannot have a classical truth value, [2] since both assuming truth and assuming falsity leads to a contradiction.

  (2) This sentence is false.

[1] This list is not exhaustive; consult, for example, [Ba81], [Ve85] and [Bl86] for more motivation.
[2] Cf. [Kr75]

- ungrammaticality or selectional restrictions: the two following sentences are also intuitively wrong, but here one cannot even dream of assigning truth or falsity, since the sentences are meaningless.

  (3) Colourless green ideas sleep furiously.

  (4) Green sleep furiously ideas colourless.

- presupposition: a sentence such as

  (5) She loves her children.

  is usually considered to be neither true nor false when 'she' in (5) is childless.

Although the evidence for partiality is overwhelming, there is a problem connected to the diversity of the arguments: since 'undefined' has a number of meanings (underdetermined, absurd, or something in between), this may corrupt the possibility of an overall partial logic covering the phenomena listed. Should we not distrust such a panacea?

Indeed we believe that we have to distinguish among different senses of 'undefined' as employed in the arguments above, since these differences are reflected in different logical behaviour. Underdeterminedness of a proposition may become irrelevant when it is combined with a true statement: for example, although 'Mary works' may lack a truth value, 'Mary works or Mary is Mary' should be counted as true. Absurdity is not removable just like that, and presuppositional undefinedness may sometimes be eliminated (when the presupposition is cancelled, for example in 'If she has children, she loves her children') but not always (cf. 'If she is a nice person, she loves her children'). Though this diversity may obstruct the emergence of a logic covering all types of partiality, it does not contradict the desirability of a partial approach. What it certainly does call for is a clear choice of the kind of undefinedness we are going to model.

In this thesis we will opt for the informational interpretation of partiality, because this is the one that is most relevant to knowledge and belief. So our main motivation stems from this area. In fact we find it difficult not to see the obvious relation between lack of truth value and lack of knowledge: with regard to available information they amount to the same. There is a slight shift of perspective: if some fact is not known to an agent, we may 'objectively' report her ignorance by saying that the proposition is unknown to her, whereas a more 'subjective' report would be that the truth of the proposition is undefined for the agent.

Apart from the direct link between knowledge and information, and the intuitions concerning this dependence, we are also confronted with less direct evidence, related to *awareness*. Some inferences or validities from classical logic badly fit real life knowledge. To wit,

(6) John knows that Mary works or does not work.

is *not* a logical truth (under the most natural sense of the word 'know'): even when we are fully informed about John's state of mind, the assertion may not be true, for John may not know Mary at all, or may not care whether she works or not. Yet, a classical semantics such as in [Hi62] and [Kr63] wrongly predicts the alleged validity. Generally speaking, partiality may cause a significant weakening of the logic, as is desirable for knowledge: unintuitive consequences can thus be removed from the logic.

Notice that in the given view, a logic does not merely consist of a set of axioms and inference rules, but contains a proper semantics as well. This licenses the very term 'partial logic', for it is the *semantics* that is incomplete, not the deductive system, nor the kind of deduction.

So, partial logic may be useful to describe logically *weaker* notions of knowledge, closer to our epistemic intuitions. But can it really cope with the psychological peculiarities of knowledge? Explicit knowledge is related to *awareness*, and chapters 6 and 7 deal with this relation. It turns out that there is no ubiquitous concept of knowledge: the epistemic logic depends on the way knowledge is used and construed. It is possible to give a general purpose system covering all weak types of logic, but one may argue that the overall logic has become almost void: virtually everything can be modelled in it, not just knowledge. Therefore we also present stronger and less general epistemic logics.

To illustrate the epistemic diversity we note that the kind of knowledge implicit in utterances (the 'added' assertoric value) requires the combination of different pieces of knowledge to account for pragmatic paradoxes such as in *saying* 'It is raining, but I don't believe it is raining' (see chapter 5), whereas an agent may also (implicitly) hold inconsistent beliefs in different frames of mind, without any chance of a mental collapse (see section 6.6). In chapter 7 we argue that partial semantics is a good candidate for describing an epistemic logic of intermediate strength, i.e. between the very weak logic of the most general models and the strong logics modelled by classical possible world semantics.

Besides appealing to intuition and to the possibility of modelling weaker logics, partiality is connected in yet another way to knowledge: the representation of knowledge in partial models can be relatively simple. The latter point is especially relevant when it comes to the representation of knowledge in computers. One striking difference with human knowledge is that computational knowledge is usually required to be stronger: in many respects, we expect an automated information system to be 'better' than ordinary human beings, i.e. not to forget, not to be confused, and, if possible, not to be incomplete.

Knowledge representation does not merely amount to storing data in a file, but also consists of providing enough structure to combine these data and derive new facts from them. Although this is usually achieved by deductive methods, we advocate a semantic approach in this book. However, it may not be easy to model weak knowledge, both in the sense of missing data and in that of limited inference.

For example, imagine the following situation. Suppose the KNMI, the Dutch Meteorological Office located in de Bilt, receives incomplete information from a weather station in, say, Amsterdam: it is either raining or snowing there. The basic translation $p \vee q$ cannot be modelled by one classical or partial propositional interpretation, since this would necessarily imply more knowledge, i.e. the model would decide which one of p and q is true (or both). So, we have to incorporate more possibilities into the model. This can be done in a possible world structure. This, at least in principle, allows a representation of more complex and accurate facts such as '*de Bilt* knows that *Amsterdam* knows that it is raining or snowing', abbreviated here as $K_b K_a(p \vee q)$. In this *modal* language we can also express queries which require some introspection of the information system, such as 'If it is raining in Amsterdam, would you know about it?'. This can be seen as a call to check whether or not $p \rightarrow K_b p$ is true in the model. Now the prospects for classical possible world models turn out to be rather limited: incomplete knowledge usually leads to a proliferation of possible worlds, if it can be modelled at all (see chapter 8). Since irrelevant aspects need not be modelled in partial semantics, there is reason to hope for a significant improvement there.

As may have incidentally become clear, a lot of partial logic is used and studied in this book. Yet, we do not wish to be narrow-minded: we will also use classical logic when more convenient, or in order to contrast the 'virtues' of the partial approach to the 'vices' of the classical one.

Overview

From the considerations above a number of more detailed issues can be obtained. In fact this thesis centres around the following themes:

- partial logic: what can be done in it and with it?
- language and knowledge: what happens to knowledge in conversations?
- knowledge and human beings: can we find more realistic accounts of human knowledge by using partial logic?
- knowledge and computers systems: can we find more efficient and adequate ways to store and retrieve information by using partial models?

The two intermediate themes are combined in one part of this book, the other themes are dealt with in separate parts. So we have the following partition, with a rough characterization of the constituting chapters:

part I explores the logical space of partial logic, in particular its expressiveness and completeness, to make a motivated choice in subsequent parts. Chapter 1 is devoted to the expressiveness of purely propositional partial logic, with a perspective akin to that of Generalized Quantifier Theory[3], in that it addresses the question which syntactic constructions correspond to which semantic constraints. Chapter 2 aims at a similar characterization of partial modal logic, although here

[3]Cf. for example [vB84c], [We84], [KM85] and [Zw83].

the similarity with first order definability in Correspondence Theory[4] is more fruitful. Chapter 3 charts the literature on validity and completeness in partial propositional logic within a unifying framework. This is extended to modal logic in Chapter 4, where proving completeness by techniques similar to the usual Henkin method for classical logic turns out to be a non-trivial enterprise.

part II deals with knowledge implicit in language use, and knowledge made explicit by human agents. In chapter 5 we assess the amount of knowledge which can be attributed to a language user in uttering assertions. In chapter 6 we will try and see to what extent human knowledge can be treated in 'classical' models (with a number of adaptations). Slightly generalizing Fagin & Halpern's logic of general awareness, we are capable of modelling virtually every logic. Since this 'most general logic' may be rather more psychological than logical, we also deal with stronger logics, more suitable for other senses of human knowledge. Much of this can be done in a more natural way in partial semantics, as demonstrated in chapter 7.

part III deals with a model-theoretic approach to knowledge bases, where chapter 8 shows the possibilities and limitations of the approach with respect to classical possible worlds models, and the concluding chapter 9 improves on this by using partial world models.

The relationship between the parts is fairly obvious: partial logic may be used to provide adequate models for human knowledge. Knowledge in its turn manifests itself in ordinary language, both in semantics and in pragmatics. Knowledge-based systems also obtain their prime motivation from human knowledge, yet in a rather idealized form. Though not implemented here, the combination of weak human knowledge and strong computational knowledge may give rise to interesting 'hybrid' logics, which can be used for building future information systems. Such an intelligent knowledge based system combines the derivation or verification procedures needed for the different kinds of knowledge, yielding a very user-friendly device. This program is well beyond the scope of this thesis — in fact it is still beyond the entire field of *artificial intelligence.*

Warning: the first part is rather technical, for reasons that hopefully are clear now. Although we would like to encourage the logically minded reader to start with part one, for less trained persons it may be wise to turn to the applications in parts two and three, and to consult the initial part afterwards, if necessary. Anyway, although there is an obvious systematic order in the chapters, a strictly linear order of reading does not seem obligatory.

[4] *Vide* [vB84b], [vB85], [vB90], etcetera

Part I

Partial logic

Introduction to part I

Even within the informational perspective on partiality underlying this thesis possibilities abound. In fact a multitude of partial logics have emerged in literature during the eighties. Our initial idea was just to give an overview of these proposals, possibly extend them with modal operators to cover knowledge, and incorporate the resulting logics in an overall system. To facilitate comparison, we restricted ourselves to the purely propositional part of these theories, focussing on two main research questions: how *expressive* is the logic, and what is the *complete* deductive system capturing the valid formulas and logical consequences. Then it turned out that, on the one hand, the field of expressiveness was still largely unexploited in the propositional case, and largely unexplored in the modal case. The issue of completeness, on the other hand, also presupposes a separate semantics. But here, the choices concerning such key notions as *rule* and *validity* were sometimes seemingly arbitrary or *ad hoc*. In other words, the attempted regimentation forced us to consider alternatives and to formulate a general framework which covers both existing proposals and alternatives.

In general our enterprise may be characterized as a *systematic exploration* of partial logic, first for propositional languages in chapters 1 and 3 and then extended to the modal language in chapters 2 and 4. This inquiry is performed in a unifying framework with enough freedom to cover the diversity observed. Although a number of the proposals in the literature do not fit into this framework, the overall theory makes clear what makes them, in our view, non-standard. Some parts of this design of a suitable framework for partial logic may well go beyond what is expected in the light of the application to knowledge representation. However, a semantics that is strictly based on private intuitions is in danger of being ill-motivated: it may disregard other qualifying possibilities.[5]

To get an idea what is at stake we first introduce a basic partial logic.

Basic partial logic

Consider the simple propositional language with a finite number of atomic propositions $p, q, \ldots$ and the ordinary connectives $\neg$ (negation, 'not'), $\wedge$ (conjunction, 'and'), $\vee$ (disjunction, 'or'), $\rightarrow$ (implication, 'if ... then') and $\leftrightarrow$ (equivalence, 'iff'). The set of well-formed formulas is defined recursively by:

- the atoms $p, q, \ldots$ are well-formed formulas (WFFs);
- if φ and ψ are WFFs, then $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, $(\varphi \rightarrow \psi)$, and $(\varphi \leftrightarrow \psi)$ are WFFs.

Formulas of this language are, for example, $p, q, \neg p, \neg q, p \wedge p, (p \wedge q) \vee \neg p$, etcetera. In a partial interpretation these formulas may be undefined. It is often easier to treat

[5]The drawback of our 'paramount' view is that the systematics causes some laborious digressions and intricate proofs, which the reader may skip at first reading.

'undefined' as if it were a third truth value, represented here by the symbol $\frac{1}{2}$.[6] Then the standard truth tables for $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ are:

$\neg$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1

$\wedge$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$\vee$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

$\rightarrow$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1	1

$\leftrightarrow$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	0	$\frac{1}{2}$	1

Expressiveness

How strong is this language? Can it express what we want to say about p and q? Recall that in classical (two-valued) logic all possible facts related to p, q, etcetera can be expressed, i.e. the logic is *functionally complete*. This does not hold for partial (three-valued) logic. For example, the proposition that is always undefined cannot be expressed in the language. Of course, this case of *functional incompleteness* can be eliminated by adding an extra connective $\star$ with is interpreted as 'undefined'. But then again, the proposition which is always true cannot be defined in the given language. Well, add another (0-place) connective $\top$ with fixed interpretation 'true'. Are we already complete? No, we still cannot express the proposition that is false if p is true, and true if p is not. So again add an operator, $\sim$, to this effect. Is the extended language functionally complete? Fortunately, it is!

This kind of game, familiar from old work in multi-valued logic, wrongly suggests that the incomplete languages are less interesting, for less expressive. The point is that perhaps we do not *want* to express all possible truth tables. Notice that before $\sim$ was added, the language had the important property of being *persistent*: if a formula is either true or false, adding extra information with respect to p and q does not result in a different truth value.

Somewhat more technically, let $\sqsubseteq$ be the (partially ordered) relation of extension of truth values such that $\frac{1}{2} \sqsubseteq 0$ and $\frac{1}{2} \sqsubseteq 1$. Moreover, let V and V' be valuations, assigning truth values to the atoms; $V \sqsubseteq V'$ means that for all atoms p: $V(p) \sqsubseteq V'(p)$. By means of the truth tables V can be extended to an interpretation $[\![\cdot]\!]_V$ of arbitrary formulas. Then persistence can be characterized with full rigour:

persistence $V \sqsubseteq V' \Rightarrow [\![\varphi]\!]_V \sqsubseteq [\![\varphi]\!]_{V'}$, for all V, V' and φ.

[6]This symbol, which can already be found in the pioneering work of Łukasiewicz, is chosen for mnemonic reasons: within the informational perspective, it is clear that $\frac{1}{2}$ can be 'resolved' by adding information, then becoming either 1 (true) or 0 (false); so it is somewhere in between, intuitively speaking.

This, of course, raises the question whether all persistent facts related to p and q can be expressed in this language. This question was answered affirmatively by Fine and Blamey.[7]

In its turn the persistence result led to a reconsideration of the intermediate languages: which conditions characterize the initial standard language and its subsequent extensions. One condition suggested by van Benthem and Langholm is *classical closure*, which says that the formula has a classical truth value when both p and q have. We will also introduce a new condition called *freedom*, which says that the truth value of every formula is undefined when p and q are. More formally, we have ($2 = \{1, 0\}$)

classical closure if $V(p) \in 2$ for all p then $[\![\varphi]\!]_V \in 2$, for all V and φ;

freedom if $V(p) = \frac{1}{2}$ for all p then $[\![\varphi]\!]_V = \frac{1}{2}$, for all V and φ.

In chapter 1 we will show that classical closure combined with freedom and persistence manages to characterize the standard language, and that freedom + persistence captures the extension with $\star$.

Four-valued semantics

There is an important alternative with respect to truth values: the truth value of a proposition may also be *overdefined*, i.e. both true and false (for instance, from different points of view). Usually, inclusion of this extra possibility does not rule out undefinedness, so it leads to a *four-valued* logic. We use the symbol 2 to incorporate 'overdefined' as a fourth truth value. The previous truth tables can be extended to this four-valued semantics, but we refrain from doing so here.[8] Since the semantics is richer, it is hardly surprising that we may consider more conditions constraining the wealth of possible truth functions. In addition to the earlier conditions we suggest the following two conditions. The first one properly generalizes classical closure, the second one means that the interpretations are invariant under switching from 'true' to not-false', and *vice versa*. Let $\tilde{1} = 1$, $\tilde{0} = 0$, $\tilde{\frac{1}{2}} = 2$, $\tilde{2} = \frac{1}{2}$, $3 = \{0, \frac{1}{2}, 1\}$ and $\tilde{3} = \{0, 1, 2\}$; moreover, let the dual $\tilde{V}$ of a valuation V be defined by $\tilde{V}(p) = \widetilde{V(p)}$ for all p.

general closure if $V(p) \in 3$ for all p then $[\![\varphi]\!]_V \in 3$, and if $V(p) \in \tilde{3}$ for all p then $[\![\varphi]\!]_V \in \tilde{3}$, for all valuations V and formulas φ;

duality preservation $[\![\varphi]\!]_{\tilde{V}} = \widetilde{[\![\varphi]\!]_V}$, for all valuations V and formulas φ.

In characterizing the classical connectives, the rôle of classical closure for the three-valued case is replaced by that of duality preservation for the four-valued case. However, both classical and general closure characterize appreciable intermediate languages.

---

[7]Cf. [Fi75a] and [Bl86], reproven here as theorem 1.3.

[8]See section 1.3 for these 'extended strong Kleene' truth tables.

There is no unanimity among 'partial' logicians whether we should have three or four truth values. The relation between 'quadrivalent' logic and classicial logic is somewhat more transparant than between 'trivalent' and classical logic (see section 2.4). [La88] and in particular [Mu89] are good examples of the use of quadrivalent approach. Most modern literature on partial logic, however, is written in the three-valued paradigm.[9]. Although our sympathy is with the three-valued approach, which seems closer to the intuitions concerning human knowledge, we have incorporated the four-valued case, since this is not only sometimes technically easier (and sometimes more difficult!), but may also be more suitable for certain computational applications. In large knowledge bases inconsistencies often occur and this should not result in a major breakdown. In the four-valued approach, inconsistencies can be verified, and thereby 'isolated'. Since in a three-valued approach (under the prevailing notions of consequence), *everything* follows from a contradiction, inconsistencies are extremely harmful for the performance of such a knowledge based system ('intelligent data base').

Varieties of validity

Now let us turn to validity, illustrated on our basic language. Typical classical validities in the standard language are $p \vee \neg p$ and $p \wedge \neg p/q$. Are they valid in partial logic as well? The answer to this question depends on how we define *validity*. For notice that $p \vee \neg p$ will not always be true, it is not valid in a 'verificational' sense. Yet, the formula is never false, so it *is* valid in a 'falsificational' sense. Both types of validity also manifest themselves for the evaluation of the argument $p \wedge \neg p/q$, but here there are additional possibilities connected to the status of the inference. One construal of inference rules is that whenever the premise is valid, so is the consequent. Since the actual premise is invalid for both types of validity (if, for example, p is true, the premise is false), the consequence relation holds in this 'absolute' construal. An alternative, 'relative' construal of inference rules is that the conclusion is true (not-false) if the premise is true (not-false, respectively). Then $p \wedge \neg p/q$ is valid with respect to *relative verification*: $p \wedge \neg p$ can never be true, so there is no counterexample to this inference. However, the inference is not valid under *relative falsification*: $p \wedge \neg p$ may be undetermined (when p is), and therefore not false, while q may be false.

In chapters 3 and 4 we give finite descriptions of the logical systems corresponding to the various notions of validity. The rules will be given in the familiar format of natural deduction. For example, the basic rules for relative verification on three-valued models are: (read $\vdash$ as 'is derivable from' and $\dashv\vdash$ as 'are derivable from each other')

[9] See for example [Ve85], [Bl86] and also [La88].

$\neg\neg\varphi \dashv\vdash \varphi$

$\neg(\varphi \wedge \psi) \dashv\vdash \neg\varphi \vee \neg\psi \qquad \neg(\varphi \vee \psi) \dashv\vdash \neg\varphi \wedge \neg\psi$

$\varphi \wedge \psi \vdash \varphi \qquad \varphi \wedge \psi \vdash \psi$

$\varphi \vdash \varphi \vee \psi \qquad \psi \vdash \varphi \vee \psi$

if $\varphi, \varrho \vdash \chi$ and $\psi, \varrho \vdash \chi$ then $\varphi \vee \psi, \varrho \vdash \chi$

if $\chi \vdash \varphi, \varrho$ and $\chi \vdash \psi, \varrho$ then $\chi \vdash \varphi \wedge \psi, \varrho$

$\varphi \wedge \neg\varphi \vdash \psi$

Apart from types of rules and validity, the semantic engine has other flexible parts: truth conditions and truth values. For example, both $\neg$ and $\sim$ may be used to interpret natural language negation, but as long as we can express 'not' in the language that satisfies the conditions, there is hardly a logical problem.[10] Really different truth conditions invoke the relation of informational extension (as used in persistence), but we consider them to be off the main track.

Extension to modal logic

Modal operators such as $\Box$ (necessity, 'must') and $\Diamond$ (possibility, 'may') can be added to the basic propositional language, yielding (propositional) modal logic. The partial models for modal logic combine the frames of classical possible world semantics with partial valuations. In partial modal semantics one usually speaks of 'situations' or 'partial worlds' rather than of possible worlds; so the valuation is supposed to manifest itself in the worlds. Similar to classical possible world semantics there is an *accessibility relation* R between situations. More precisely, a partial modal model M is a triple $\langle S, R, V\rangle$, where $R \subseteq S \times S$ and V is a partial function mapping atom–situation pairs into $\{0, 1\}$.

The truth table format does not suit this type of semantics. Instead we use the partial truth relation $\models\!\!\!\equiv$ and the falsity relation $\equiv\!\!\!\dashv$.[11] We may recast the truth tables for the connectives in the new format and incorporate the truth conditions for the modal operators $\Box$ and $\Diamond$ as: (the model $\langle S, R, V\rangle$ is fixed in these clauses)

$$\begin{array}{ll}
s \models\!\!\!\equiv p \Leftrightarrow V(p,s) = 1 \ (\forall p \in Prop) & s \equiv\!\!\!\dashv p \Leftrightarrow V(p,s) = 0 \ (\forall p \in Prop) \\
s \models\!\!\!\equiv \neg\alpha \Leftrightarrow s \equiv\!\!\!\dashv \alpha & s \equiv\!\!\!\dashv \neg\alpha \Leftrightarrow s \models\!\!\!\equiv \alpha \\
s \models\!\!\!\equiv \alpha \wedge \beta \Leftrightarrow s \models\!\!\!\equiv \alpha \ \& \ s \models\!\!\!\equiv \beta & s \equiv\!\!\!\dashv \alpha \wedge \beta \Leftrightarrow s \equiv\!\!\!\dashv \alpha \ or \ s \equiv\!\!\!\dashv \beta \\
s \models\!\!\!\equiv \alpha \vee \beta \Leftrightarrow s \models\!\!\!\equiv \alpha \ or \ s \models\!\!\!\equiv \beta & s \equiv\!\!\!\dashv \alpha \vee \beta \Leftrightarrow s \equiv\!\!\!\dashv \alpha \ \& \ s \equiv\!\!\!\dashv \beta \\
s \models\!\!\!\equiv \Box\varphi \Leftrightarrow \forall t : sRt \Rightarrow t \models\!\!\!\equiv \varphi & s \equiv\!\!\!\dashv \Box\varphi \Leftrightarrow \exists t : sRt \ \& \ t \equiv\!\!\!\dashv \varphi \\
s \models\!\!\!\equiv \Diamond\varphi \Leftrightarrow \exists t : sRt \ \& \ t \models\!\!\!\equiv \varphi & s \equiv\!\!\!\dashv \Diamond\varphi \Leftrightarrow \forall t : sRt \Rightarrow t \equiv\!\!\!\dashv \varphi
\end{array}$$

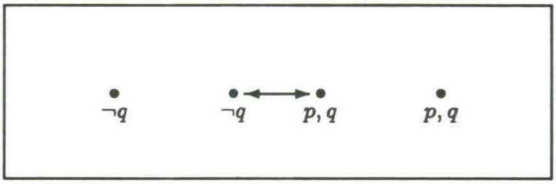
[10] There may be a problem from the standpoint of natural language semantics in formulating a compositional translation procedure from syntax into logic, but this does not concern us in this thesis.

[11] These symbols, sometimes disrespectfully called 'rakes', are convenient since they express the partial character of truth and falsity, without indices messing up the notation. The distinction between classical and partial truth is not merely cognitive, but also serves a technical goal, especially in a hybrid style of semantics in which both types of truth relations are used within a single model (cf. chapter 7). $\models\!\!\!\equiv$ was inspired by Kamp's notation for strong consequence. Other common symbols for $\models\!\!\!\equiv$ and $\equiv\!\!\!\dashv$ are: $\models$ and $\dashv$, $\models^+$ and $\models^-$, or $\models_T$ and $\models_F$.

This model non-falsifies $K\neg p \vee Kq \vee K\neg q$, which is not an F–consequence of Kp. Similarly, dropping 'p' in the right-hand world of the middle component admits the non-consequence $\neg Kp \vee Kq \vee K\neg q$. In fact, as the reader may check (warning: this is tedious labour), none of the occurences of 'p', 'q' and '$\neg q$' may be omitted without loss of characterization.

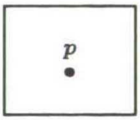
These examples suggest that minimal total F–miniatures are unique, up to isomorphism. The examples and the previous proposition may also suggest the generalization that classical (**S5**)- miniatures for some set of data D are always F–miniatures for D as well. Though tempting, the latter is not true.

Example 9.3 *Consider the data $D = \{Kp, K(p \to q)\}$. A total model that verifies D will also verify Kq. Consequently, essentially the only classical* **S5** *miniature for D will be the singleton model verifying both p and q. But Kq is not an F–consequence of D, roughly because* **S5*** *does not contain* Modus Ponens. *So the singleton model is not an F–miniature for D. A small F–miniature for D is:*



Another example of the incongruity of classical and V–miniatures may be more transparant. The point is that for inconsistent data, F–consequence and **S5**-consequence diverge widely.

Example 9.4 *For the data $\{Kp, Kq, K\neg q\}$ the minimal F–miniature is:*



As in the previous example, there is no total F–miniature for this set of data: a total model that non-falsifies q and $\neg q$ in each world should verify q and $\neg q$ in each world, thus has to be the empty model. But the empty model also non-falsifies $K\neg p$, which does not follow from the data in **S5***.

From these comparisons between (partial and total) F–miniatures and (total) classical miniatures some generalizations are induced:

- A total F–miniature for D is also a classical miniature for D.
- If D has a minimal F–miniature that is partial, it has no total F–miniature.

An early combination of partial and modal logic is [Se67], where the underlying propositional logic is characterized by the 'weak Kleene' truth tables, cf. [Bo37] and [Kl52]. The transition from worlds to situations was made independently by several authors, most explicitly by [Hu81] and [Ba81]. Barwise's theory, covering the semantics of perception verbs, has been formalized and studied to considerable depth in [Ka83].

Chapter 1

Propositional definability

This chapter concerns the expressiveness of propositional languages with respect to partial semantics. After providing some background for the subject of functional completeness and partial semantics, we define the key notion of *definability* of a set of (partial) truth functions by means of a set of connectives. Interesting sets of truth functions are those determined by general conditions; we will restrict ourselves to conditions with a very natural ring.

Many conditions such as *persistence* and *classical closure*[1] are definable by certain connectives. Some isolated definability results were already known, many others were not. This chapter approaches the subject in a more systematic way, and provides a considerable number of new definability results, some for 3–valued models and some for 4–valued models.

Our general strategy is to start with no condition at all and give a general 'functional completeness' theorem. Then we add conditions, one after the other, thus successively narrowing down the set of defining connectives.

After proving and reproving several old and new definability results we turn to an issue which has apparently been neglected so far: the relative and absolute *strength* of the conditions. More precisely, how many functions of certain arity are allowed by some combination of conditions? After a summary of the main results, there is an appendix containing a more detailed mathematical account of definability matters and a comparison to existing literature concerning conditions and connectives.

1.1 Introduction

To explore the field of partial logic, we first have to inquire into its expressiveness: which classes of partial models can be defined by which sets of formulas? This chapter focusses on questions of *propositional* definability, in relation to conditions on the set of (3– or 4–valued) truth functions. Some isolated results are known in this area, whereas other plausible questions have not been answered so far. The aim of this chapter is simply to partly fill the gap.

[1]These conditions are informally described in the introduction to part I.

1.1.1 Definability in classical logic

In classical propositional logic one easily obtains so-called functional completeness for the usual set of connectives, or a proper subset such as $\{\neg, \wedge, \}$: every bivalent truth function can be expressed in a logical formula, using only the given connectives propositional variables.

Theorem 1.1 *The bivalent truth functions are definable by* $\neg$ *and* $\wedge$.

Proof: Let $f : 2^n \longrightarrow 2$ be a classical n-place truth function ($2 = \{0,1\}$). The behaviour of f on $\vec{x} \in 2^n$ is described by a conjunction of literals: when $f(\vec{x}) = 1$ the literal is p_i if $x_i = 1$ and $\neg p_i$ if $x_i = 0$; when $f(\vec{x}) = 0$ both literals p_i and $\neg p_i$ occur in the conjunction. Then the formula χ_f, which characterizes f, is the disjunction of the conjunctive subformulas for all possible arguments $\vec{x}$. So formally:

$$\begin{aligned}
\chi_{\vec{x},i} &= \begin{cases} p_i & \text{if } x_i = 1 \text{ and } f(\vec{x}) = 1 \\ \neg p_i & \text{if } x_i = 0 \text{ and } f(\vec{x}) = 1 \\ p_i \wedge \neg p_i & \text{if } f(\vec{x}) = 0 \end{cases} \\
\chi_f &= \bigvee_{\vec{x}} \bigwedge_{i=1}^{i=n} \chi_{\vec{x},i}
\end{aligned}$$

The truth table of χ_f coincides with f (i.e. $[\![\chi_f]\!]_{\vec{x}} = f(\vec{x})$ for each valuation $[\![\cdot]\!]_{\vec{x}}$ that assigns x_i to p_i), since $[\![\bigwedge_i \chi_{\vec{y},i}]\!]_{\vec{x}} = f(\vec{x})$ if $\vec{x} = \vec{y}$, and 0 otherwise.
Finally disjunction can be eliminated by the equivalence of $\varphi \vee \psi$ and $\neg(\neg\varphi \wedge \neg\psi)$. ■

1.1.2 Definability in partial logic

For partial logic, where at least the value $\frac{1}{2}$ ('undefined') is added to the classical truth values 1 ('true') and 0 ('false'), functional completeness no longer holds: three-valued functions (from 3^n to $3 = \{0, \frac{1}{2}, 1\}$) [2] cannot in general be described by the stock of classical connectives, since for example the constant truth functions are not definable in this way. So, the function f such that $f(x) = 1$ where $x = 1, \frac{1}{2}$ or 0, is not definable by means of a formula using only p, $\neg$ and $\wedge$, as will become clear in the course of this chapter. Of course the same holds for the four-valued case which contains the extra truth value 2 ('overdefined').

One possible reaction to the obtained incompleteness is to add new connectives that restore functional completeness. This can be done in many ways.[3] Though extensionally equivalent to other complete systems, we will favour some systems which we believe to be more natural than others. Our favourite systems are perfectly equipped for the weak, yet flexible propositional logic involved in knowledge and belief. [4]

---

[2] So, the symbol '3' may refer to the ordinary ordinal number 3 and to a set of truth values. Also, '2' may indicate a set of truth values as in the proof of theorem 1.1, or a truth value itself; the latter '2' is the algebraic counterpart of the set (of underlying truth values) $\{0,1\}$, the former of the set $\{\{0\},\{1\}\}$. Likewise the truth value 1 has a set-theoretical counterpart $\{1\}$, etcetera. Though this may seem rather confusing, all ambiguities should be resolved by the context.

[3] See [Ro77] for general properties of functionally complete systems.

[4] See chapters 5–9 of this book.

Another possible reaction to incompleteness is to accept it as a consequence of allowing truth value gaps and overlaps, and to search for general constraints that single out classes of functions which can be defined by a set of connectives. Then definability questions appear in two related forms:

- Given some set of connectives, which semantic constraints characterize them?
- Given some set of semantic constraints, which bunch of connectives can be used to define precisely the functions fulfilling this constraint?

We shall discuss definability matters both from the perspective of natural conditions and from that of intuitively plausible connectives. Since definability results can be obtained from both perspectives, we do not employ a systematic division between them in the next section.

Before we go on, we have to be more precise about some key notions.[5] Definability is one way of relating the syntactic and the semantic side of a logic. Assume there are k truth values.[6] For some n, $\vec{x}$ is short for $\langle x_1, \ldots, x_n\rangle$ (with $x_i \in k$), and k^n is the set of such $\vec{x}$'s. Think of n as being the finite number of propositional atoms in *Prop* and $\vec{x}$ as essentially a valuation.[7] k^{k^n} will be the set of n-ary k-valued truth functions. Then the set of k-valued truth functions is called the (k-valued) function space.

Definition 1.1 (function space)
F_k, *the* **function space** *for the k-valued case, denotes* $\bigcup_{n\in\omega} k^{k^n}$.

The syntactic side of definability is, of course, the logical language, i.e. the set of well-formed formulas. Let C be a set of connectives (possibly 0-place). Unless stated differently, we assume C to contain at least $\neg$ and $\wedge$.

Definition 1.2 (propositional language)
The set of well-formed formulas $\mathcal{L}_C(Prop)$, with logical constants in C and using the atoms of Prop, is the smallest set containing Prop which is closed under the operations of C.

In other words, for given *Prop* and C, $\mathcal{L}_C(Prop)$ is recursively defined by[8]

1. $Prop \subseteq \mathcal{L}_C(Prop)$,
2. if $c \in C$ is an n-ary operation and $\{\varphi_1, \ldots, \varphi_n\} \subset \mathcal{L}_C(Prop)$, then $c\varphi_1 \cdots \varphi_n \in \mathcal{L}_C(Prop)$,
3. no other elements occur in $\mathcal{L}_C(Prop)$ than produced by 1 and 2.

---

[5] A number of technicalities such as a definition of the notion of a *closed class* of truth functions have been transferred to the appendix.

[6] In this book, k is either 2, 3, or 4.

[7] This is related to the more usual form given in the general introduction by $V(p_i) = x_i$, and similarly $[\![\varphi]\!]_V = [\![\varphi]\!]_{\vec{x}}$.

[8] We leave out *Prop* or C from $\mathcal{L}_C(Prop)$ when irrelevant or clear from context.

For two-place operators such as $\wedge$ and $\vee$, we will use the more common infix notation in parentheses, for example $p \vee (q \wedge r)$ instead of $\vee p \wedge qr$.

Instead of merely defining functional completeness, we want to have a more general notion of expressiveness at hand: definability of a restricted set of truth functions by means of some set of connectives or other. Think of C as a (combined) condition singling out truth functions and let $\mathsf{C}^{(n)} = \mathsf{C} \cap k^{k^n}$.

Definition 1.3 (propositional definability)
$\mathsf{C} \subseteq \mathsf{F}_k$ is definable by a set of connectives C iff (i) for each $f \in \mathsf{C}^{(n)}$ there is a $\varphi \in \mathcal{L}_C\{p_1, \ldots, p_n\}$ such that $[\![\varphi]\!]_{\vec{x}} = f(\vec{x})$ for all $\vec{x} \in k^n$, where $[\![p_i]\!]_{\vec{x}} = x_i$; and (ii) for all $\varphi \in \mathcal{L}_C\{p_1, \ldots, p_n\}$ the mapping $f : \vec{x} \mapsto [\![\varphi]\!]_{\vec{x}}$ is in $\mathsf{C}^{(n)}$.

1.2 Classes of three-valued functions

A functionally complete system for F_k in general was already given in [Po21].[9] Yet, in line with what follows here, we want to pay attention to particular systems for F_3 and F_4 which we believe to be more intuitive than the one featuring in Post's general result. Focussing on the 3-valued case first, our favourite system uses 0–place $\star$ with constant interpretation $\frac{1}{2}$, strong Kleene $\wedge$, and two negations, a standard one ($\neg$) and a non-standard one ($\sim$), attributed to Bochvar who is supposed to have named it 'external negation'.[10] The truth tables for the negations are:

	1	$\frac{1}{2}$	0
$\neg$	0	$\frac{1}{2}$	1

	1	$\frac{1}{2}$	0
$\sim$	0	1	1

$\sim$ is discussed in [FH*87] and [La88]; in the former it is called 'weak negation', in the latter 'exclusion negation'; yet another (folk) name is 'Nordic negation', for obvious reasons. $\sim$ gives a classical flavour to partial systems, which has interesting consequences for the set of tautologies.

From the truth tables of $\neg$ and $\wedge$, those for $\vee$, $\rightarrow$ and $\leftrightarrow$ can be derived by the usual definitions: $\varphi \vee \psi = \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi = \neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi = (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. Since $\wedge$ and $\vee$ both are particularly useful in definability proofs, we repeat their truth tables below:

$\wedge$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0
0	0	0	0

$\vee$	1	$\frac{1}{2}$	0
1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

[9] For the initial part of the natural numbers $\{0, 1, \ldots, k-1\}$ with the usual ordering and addition, Post's system consists of the *generalized conjunction* $\wedge_k$ and *cyclical negation* $\neg_k$ which are semantically characterized by: $x \wedge_k y = \min\{x, y\}$ and $\neg_k x \equiv x + 1 (\bmod k)$.

[10] The attribution of both Urquhart [Ur86, p.75] and Langholm [La88, p.17], possibly following Rescher [Re69, pp.31,32], to Bochvar is not entirely correct: [Bo37, E. tr. pp.91,93] gives a 'formal external denial' with the intention of formalizing *being false* and which produces a different truth table, viz. one that has 0 when the formula is undefined; it can be defined in the [FH*87]-format by e.g. $\sim\sim\neg$. The negation corresponding to $\sim$ is nameless, and captures the intuition of *being not true*.

By a direct proof (cf. theorem 1.6 on page 25), a reduction to another system known to be complete (for example, to the Post system by $\neg_3 p = (p \wedge \star) \vee (\sim p \wedge \sim\neg p)$ and $\wedge_3 = \wedge$), or an application of the Słupecki completeness criterion[11], it can easily be shown that

Theorem 1.2
Every trivalent truth function is definable by $\{\neg, \wedge, \star, \sim\}$. [12]

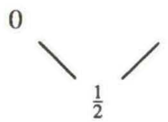
Notice moreover that the system presented in this theorem is independent (non-redundant): every connective it contains is necessary. This follows from the fact that subsystem such as $\{\neg, \wedge\}$ are incomplete.

As was indicated before the somewhat isolated and seemingly dead branch of many-valued logic involved in definability revived as soon as one got interested in *sub*classes of the entire function space, and, correspondingly, certain conditions on admissible operators.

One of the first logically relevant results in this direction concerns Belnap's notion of *persistence* (or, *monotonicity*):

PERS f is *persistent* iff for all $\vec{x}, \vec{y}$: $\vec{x} \sqsubseteq \vec{y} \Rightarrow f(\vec{x}) \sqsubseteq f(\vec{y})$,

where $f : 3^n \longrightarrow 3$, $\vec{x} \sqsubseteq \vec{y}$ means that $x_i \sqsubseteq y_i$ for every $i = 1, \ldots, n$, and $\sqsubseteq$ is the approximation relation which is the partial order such that $\frac{1}{2} \sqsubseteq 0$, $\frac{1}{2} \sqsubseteq 1$. The simple approximation semi-lattice $\langle 3, \sqsubseteq \rangle$ that this gives rise to is depicted below:



Notice that this semi-lattice is closed with respect to the *meet* operation $\sqcap$, but not with respect to the join $\sqcup$. Persistence is an important notion, which can be motivated independently, cf. [Bl86].

By adding some connectives to the classical ones ($\neg$ and $\wedge$), for example the constants $\top$ ('*verum*', interpreted as 1) and $\star$ ('*ignoratio*'), we arrive at what is presumably the first definability result of this sort:

Theorem 1.3 (Fine/Blamey[13])
The persistent trivalent truth functions are definable by $\{\neg, \wedge, \top, \star\}$.

---

[11]The criterion of [Sł39] says that a subset of F_k is complete for F_k iff it contains all unary functions and at least one surjection which is not equivalent to a unary function, where f is equivalent to unary g means that there is an i such that for all $\vec{x}$: $f(\vec{x}) = g(x_i)$, i.e. f depends on only one of its arguments.

[12]The theorem in [La88, p.29] amounts to the same, despite its very different appearance and proof. His language $\mathbf{l}_{\sim,\star}$ also uses the symbols $\neg$, $\vee$ and $\top$; of course $\vee$ and $\wedge$ are interdefinable modulo $\neg$, and $\top$ is redundant: $[\![\top]\!] = [\![\sim\star]\!]$.

Proof: a simple version of the proof runs as follows. First notice that the interpretations of atoms, standard connectives and constants are persistent, cf. proposition 3.4. Next define $\vee$ and $\curlyvee\!\curlyvee$ by:

$$\varphi \vee \psi = \neg(\neg\varphi \wedge \neg\psi)$$
$$\varphi \curlyvee\!\curlyvee \psi = (\star \wedge \varphi) \vee (\varphi \wedge \psi) \vee (\psi \wedge \star)$$

Then for a given $f : 3^n \longrightarrow 3$ define $\varphi_{\vec{x},i} = \begin{cases} p_i & \text{if } x_i = 1 \\ \neg p_i & \text{if } x_i = 0 \end{cases}$.
If $f = 1$, let $\varphi_f = \top$ and if $f = 0$, let $\varphi_f = \neg\top$; otherwise

$$\varphi_f = \left(\bigvee\nolimits_{\vec{x}}^{f(\vec{x})=1} \bigwedge\nolimits_i^{x_i \neq \frac{1}{2}} \varphi_{\vec{x},i}\right) \curlyvee\!\curlyvee \left(\neg \bigvee\nolimits_{\vec{x}}^{f(\vec{x})=0} \bigwedge\nolimits_i^{x_i \neq \frac{1}{2}} \varphi_{\vec{x},i}\right),$$

where, for each $\vec{x}$, $\bigwedge$ is the finite conjunction of the $\varphi_{\vec{x},i}$ such that $x_i \neq \frac{1}{2}$, and $\bigvee$, similarly, the finite disjunction of these conjunctions for which $\vec{x}$ satisfies the displayed condition. Then $[\![\varphi_f]\!]_{\vec{x}} = f(\vec{x})$.[14] ∎

Theorem 1.3 provides an answer to the question which set of connectives defines persistence,[15] and so illustrates the second research question from the introduction. Alternatively, we can start at the other end: given a set of connectives look for the conditions characterizing them.

An evident question of the latter type concerns the suitable constraints for the classical connectives, i.e. those definable by, for example, $\neg$ and $\wedge$. From theorem 1.3 it follows that PERS has to be one of the constraints, but also that there must be others eliminating $\top$ and $\star$, and our task is to find them. An important move in this direction was made independently by (among others) Albert Visser and Johan van Benthem who propose the obvious condition of what we will call *classical closure*:[16]

CCLOS f is *classically closed* iff $f[2^n] \subseteq 2$, where $2 = \{0, 1\}$.

The condition excludes $\star$: for $n = 0$ the only 0-place functions are those corresponding to $\top$ and $\bot$, since $2^0 = 3^0 = \{\emptyset\}$. Combination of PERS and CCLOS leads to the following theorem:

Theorem 1.4 (Van Benthem/Langholm)
Every classically closed persistent trivalent truth function can be defined by a formula using $\{\neg, \wedge, \top\}$.

[13]This result is usually attributed to Blamey [Bl86, p.36,37], but can already be found in [Fi75a, p.288], who uses the phrase 'stability' instead of persistence. The proof is Blamey's, though.

[14]Notice that this proof uses a *partial* (!) definition of $\varphi_{\vec{x},i}$; we can turn it into a full one by adding $\varphi_{\vec{x},i} = \top$ if $x_i = \frac{1}{2}$.

[15]Of course this set is not unique; for example, $\{\rightarrow, \bot, \star\}$ defines PERS_3 too; in fact we can do better than this and give a complete system of two connectives, but in [Th90c, appendix] it is shown that no single connective can define persistence. But already the two connectives system fails to have the naturalness of the text system.

[16]This condition is equivalent to what is called 'refinability' in [Ja91a] (but cf. the appendix for different notions of refinability): f is *refinable* iff $f(\vec{x}) = \frac{1}{2}$ implies that there exist $\vec{y}, \vec{z} \sqsupseteq \vec{x}$ such that $f(\vec{y}) = 1$ *or* $f(\vec{z}) = 0$.

Proof: from now on we will mostly skip the 'correctness' part of the definability theorems, which can be shown by a straightforward induction. For the 'completeness' part, employ the construction of [vB88], which is essentially the following.[17] Let $f : 3^n \longrightarrow 3$ be a closed persistent function. The characterizing formula φ_f takes the form

$$\varphi_f = \bigvee_{\vec{x}} \bigwedge_{i=1}^{i=n} \varphi_{\vec{x},i}.$$

The contributions $\varphi_{\vec{x},i}$ are defined by:

- if $f(\vec{x}) = 1$ then $\varphi_{\vec{x},i} = \begin{cases} p_i & \text{if } x_i = 1 \\ \neg p_i & \text{if } x_i = 0 \\ \top & \text{if } x_i = \frac{1}{2} \end{cases}$
- if $f(\vec{x}) = \frac{1}{2}$ then $\varphi_{\vec{x},i} = \begin{cases} p_i & \text{if } x_i = 1 \\ \neg p_i & \text{if } x_i = 0 \\ p_i \wedge \neg p_i & \text{if } x_i = \frac{1}{2} \end{cases}$
- if $f(\vec{x}) = 0$ then $\varphi_{\vec{x},i} = \bot = \neg\top$

Then an intricate argument shows that for the valuation $[\![\cdot]\!]_{\vec{x}}$ that assigns x_i to p_i: $[\![\varphi_f]\!]_{\vec{x}} = f(\vec{x})$. ∎

Addopting the other perspective on definability questions, we note that PERS+CCLOS can indeed be defined by other systems, such as $\{\rightarrow, \bot\}$.[18]

To see where we are, let us give a recapitulation of the main results. Starting with no condition at all we noticed definability by $\neg, \wedge, \top, \star$ and $\sim$ ($\top$ has been added to enable comparison). Then subsequent addition of the conditions *persistence* and *classical closure* amounted to elimination of the connectives $\sim$ and $\star$, respectively. To characterize the classical connectives a final step is needed: how to exclude $\top$ and those functions definable by it, such as $\bot = \neg\top$ and $p \wedge \neg\top$. The proposed condition is dubbed *freedom*, since when a function does not satisfy the requirement the whole function is determined, *modulo* persistence.

FREE f is called *free* iff $f(\vec{\frac{1}{2}}) = \frac{1}{2}$.

where $\vec{\frac{1}{2}} = \overbrace{\langle\frac{1}{2}, \ldots, \frac{1}{2}\rangle}^{n\times}$. So adding *freedom*[19] to persistence eliminates essentially two functions, viz. for each arity n the constant functions **0** and **1**. In the spirit of information and partiality *freedom* is a very natural condition. The premeditated definability result is now embodied by:

Theorem 1.5
Every free classically closed persistent trivalent truth function is definable by $\{\neg, \wedge\}$.

[17] In order to diversify the presentation we give a full definition where [vB88] gives a partial one. The structure of Tore Langholm's proof is more complex.

[18] But again *not* by a single connective, see [Th90c, appendix].

[19] Notice this amounts to closure with respect to $\{\frac{1}{2}\}$.

Proof: an adaptation of the proof for the previous theorem suffices.

First notice that all the (classical) connectives with standard interpretations are free: a (classical) truth value is only obtained when at least one of atoms has a classical truth value.

Second, by *freedom* there is an $\vec{x}$ such that $f(\vec{x}) \neq 0$, viz. $\vec{x} = \vec{\frac{1}{2}}$; therefore, all disjuncts for $\vec{x}$ such that $f(\vec{x}) = 0$ can be omitted by logical equivalence from φ_f, without making the formula empty.

Third, conjuncts of the form $\top$ can also be left out; the only case where such a deletion would be troublesome is where all x_i are $\frac{1}{2}$ and $f(\vec{x}) = 1$; but again this case is excluded because of the fact that f is free. In all, a *partial* definition of $\varphi_{\vec{x},i}$ works where the last item and the third case of the first item in the earlier definition are left out.

So, for example the function of the 'weak Kleene conjunction' described by the matrix

	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	0	$\frac{1}{2}$	0

is characterized by the formula $(p \wedge q) \vee (p \wedge q \wedge \neg q) \vee (p \wedge \neg p \wedge q) \vee (p \wedge \neg p \wedge q \wedge \neg q) \vee (p \wedge \neg p \wedge \neg q) \vee (\neg p \wedge q \wedge \neg q)$ which is equivalent to $(p \wedge q) \vee (p \wedge \neg p) \vee (q \wedge \neg q)$ ■

This completes what we consider to be the main trail of this section. However, given the conditions that were introduced before, we may wonder by which connectives they are definable, both in isolation and in combination. To begin with, recall a result presented in [vB88]. Which set of connectives defines precisely the classically closed functions? To this end we reconsider the weak negation $\sim$, which is clearly not persistent, but it is classically closed and its addition to the classical connectives guarantees functional completeness relative to CCLOS:[20] the trivalent classically closed truth function are definable by means of $\neg, \wedge, \sim$. [21]

A complementary combination (with respect to theorem 1.5) deals with FREE and PERS. The relevant characterization can be simply obtained by an adaptation of Blamey's proof: the free persistent truth functions are definable by $\{\neg, \wedge, \star\}$.[22]

The latter fact has an interesting consequence. Instead of laborious arguments showing functional completeness of some combination of conditions by a set of connectives, one might want a simple and general test predicting completeness. There is some reason to doubt the existence of such a simple test. To illustrate the point notice that the Słupecki criterion fails: all unary FREE functions are also persistent, hence definable by $\neg, \wedge, \star$, which operators are free, one of them being a non-reducible two-place function ($\wedge$). Yet several binary free functions are not persistent, thus cannot be defined by these operators alone.

[20]CCLOS can also be defined by a smaller set of connectives, such as Łukasiewicz's Ł$_3$ (see the appendix to this chapter), or even by a single connective (see [Th90c, p.39]).

[21]The proof "is an easy matter of recording the truth table as in the classical two-valued case" [vB88, p.85]. In [Th90c, pp.15,16] there is an elaborated version of this proof, cf. the proof of theorem 1.11.

[22][Th90c, p.16]

1.3 Classes of four-valued functions

Again the story starts with a familiar result: the functional completeness of F_4. Recall that the functional completeness of one particular system was already shown in [Po21]. Once more, the connectives occurring in the following theorems are preferable, since they also apply to certain conditions to be dealt with in this section. The first new connective involved here is the 0-place $\sharp$ with constant interpretation 2 (*overdefined*); $\star$ is interpreted as $\frac{1}{2}$ as before. The extended strong Kleene truth tables for $\neg$, $\wedge$ and $\vee$ are displayed below. To motivate the extension of the interpretation of $\sim$ to the 4-valued case, notice that the basic intuition behind $\sim$ is that $\sim\varphi$ is verified (supported) whenever φ is not, and that $\sim\varphi$ is falsified (rejected) when φ is verified.[23]

$\neg$	
1	0
$\frac{1}{2}$	$\frac{1}{2}$
0	1
2	2

$\wedge$	1	$\frac{1}{2}$	0	2
1	1	$\frac{1}{2}$	0	2
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
0	0	0	0	0
2	2	0	0	2

$\vee$	1	$\frac{1}{2}$	0	2
1	1	1	1	1
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
0	1	$\frac{1}{2}$	0	2
2	1	1	2	2

$\sim$	
1	0
$\frac{1}{2}$	1
0	1
2	0

Theorem 1.6 (Langholm)
Every quadrivalent truth function is definable by $\{\neg, \wedge, \star, \sim, \sharp\}$.

Proof:[24] Let f be an arbitrary four-valued n-place truth function. Then the formula characterizing f is built from subformulas *encoding* $\vec{x}$ and $f(\vec{x})$ for each $\vec{x}$, where such a subformula is only non-zero for the interpretation corresponding to $\vec{x}$. First we define some auxiliary formulas that characterize some 'most specific' unary truth functions, as illustrated by their truth tables:

p	$\chi_{1000}(p) =$ $\neg\sim p \wedge \sim\neg p$	$\chi_{0100}(p) =$ $\sim p \wedge \sim\neg p$	$\chi_{0010}(p) =$ $\sim p \wedge \neg\sim\neg p$	$\chi_{0001}(p) =$ $\neg\sim p \wedge \neg\sim\neg p$
1	1	0	0	0
$\frac{1}{2}$	0	1	0	0
0	0	0	1	0
2	0	0	0	1

Next let

$$\chi_{\vec{x},i} = \begin{cases} \chi_{1000}(p_i) & \text{if } x_i = 1 \\ \chi_{0100}(p_i) & \text{if } x_i = \frac{1}{2} \\ \chi_{0010}(p_i) & \text{if } x_i = 0 \\ \chi_{0001}(p_i) & \text{if } x_i = 2 \end{cases}$$

The conjuncts $\varphi_{\vec{x},i}$ are then defined by distinguishing the following cases:

- if $f(\vec{x}) = 1$ then $\varphi_{\vec{x},i} = \chi_{\vec{x},i}$
- if $f(\vec{x}) = \frac{1}{2}$ then $\varphi_{\vec{x},i} = \chi_{\vec{x},i} \wedge \star$
- if $f(\vec{x}) = 0$ then $\varphi_{\vec{x},i} = \star \wedge \sharp$

[23] I.e. formally: $0 \sqsubseteq [\![\sim\varphi]\!] \Leftrightarrow 1 \sqsubseteq [\![\varphi]\!]$ and $1 \sqsubseteq [\![\sim\varphi]\!] \Leftrightarrow 1 \not\sqsubseteq [\![\varphi]\!]$.

[24] Where [La88, pp.27,28] uses a reduction to classical models and [Mu89, p.123] gives a nice inductive proof, we prefer a more direct argument for expository reasons.

- if $f(\vec{x}) = 2$ then $\varphi_{\vec{x},i} = \chi_{\vec{x},i} \wedge \sharp$

The formula corresponding to f is

$$\varphi = \bigvee_{\vec{x}} \bigwedge_{i=1}^{i=n} \varphi_{\vec{x},i}.$$

From the truth tables it follows that if $\varphi_{\vec{y}} = \bigwedge_{i=1}^{i=n} \varphi_{\vec{y},i})$ $\vec{y} \neq \vec{x}$, then $[\![\varphi_{\vec{y}}]\!]_{\vec{x}} = 0$ if $\vec{y} \neq \vec{x}$, and $[\![\varphi_{f,\vec{x}}]\!]_{\vec{x}} = f(\vec{x})$. So $[\![\varphi]\!]_{\vec{x}} = f(\vec{x})$. ■

As in the previous section we want to cut down the number of defining connectives by imposing more and more conditions, until we have reached the classical connectives (i.e. $\neg$ and $\wedge$). Once again, the natural condition of *persistence* is a good starting point. For F_4, persistence is a condition on functions $f : 4^n \longrightarrow 4 = \{0, \frac{1}{2}, 1, 2\}$ which preserve the approximation relation $\sqsubseteq$. The graph of the approximation lattice $\langle 4, \sqsubseteq \rangle$ is displayed below: (recall that $\sqsubseteq$ amounts to going upwards in the diagram)

$$\begin{tikzcd}[column sep=small] & 2 & \\ 0 \arrow[ur, dash] & & 1 \arrow[ul, dash] \\ & \frac{1}{2} \arrow[ul, dash] \arrow[ur, dash] & \end{tikzcd}$$

Reinhard Muskens[25] has extended the Fine/Blamey persistence theorem to four-valued functions by incorporating the constant $\sharp$.[26]

Theorem 1.7 (Muskens)
The persistent quadrivalent truth functions are definable by $\{\neg, \wedge, \star, \sharp\}$.

It may seem at this point that one may transfer definability results from the three-valued case to the four-valued case by simply adding the operator $\sharp$ to the set of connectives. In general there is no guarantee that such a strategy will always succeed. In effect, the addition of $\sharp$ to the connectives defining CCLOS will not do: $\sharp$ is not classically closed itself.[27] One may object that $\sharp$ should be put on a par with $\star$. Though this seems intuitively right and will indeed be employed in the sequel, we notice that persistence and classical closure do not jointly characterize $\{\neg, \wedge, \top\}$: the function f given by

	1	$\frac{1}{2}$	0	2
f	0	$\frac{1}{2}$	0	0

is persistent and classically closed, but not definable by $\neg, \wedge, \top$. Whence we cannot transfer van Benthem's result (theorem 1.4) to F_4 that easily. Let us therefore look for

[25] *Vide* [Mu89, pp.46,124/5]; Muskens' proof resembles Blamey's (cf. theorem 1.3).
[26] Notice that $\top$ is now redundant: $\top = \neg(\star \wedge \sharp)$.
[27] For those who prefer a counterexample of non-zero arity: $p \wedge \sharp$ is not CCLOS.

another condition characterizing the functions definable by means of $\{\neg, \wedge, \top\}$. One property that $\neg$, $\wedge$ and $\top$ share, but the above f fails to have is *duality preservation*.[28]

DUAL f duality preserving iff $f(\tilde{\vec{x}}) = \widetilde{f(\vec{x})}$ for all $\vec{x}$.

where $\tilde{1} = 1, \tilde{0} = 0, \tilde{\frac{1}{2}} = 2, \tilde{2} = \frac{1}{2}$,[29] and $\tilde{\vec{x}} = \langle \tilde{x}_1, \ldots, \tilde{x}_n \rangle$ if $\vec{x} = \langle x_1, \ldots, x_n \rangle$. Now in effect the addition of DUAL makes CCLOS redundant:

Proposition 1.1 *DUAL $\Rightarrow$ CCLOS*

Proof: We give an indirect proof. Assume that there is an $f \in \mathsf{F}_4$ that is duality preserving but is not classically closed; i.e. there is an $\vec{x} \in 2^n$ such that $f(\vec{x}) \notin 2$. So $\tilde{\vec{x}} = \vec{x}$ and $\widetilde{f(\vec{x})} \neq f(\vec{x})$ which contradicts DUAL. ■

Unlike CCLOS, the addition of DUAL to PERS captures $\{\neg, \wedge, \top\}$:

Theorem 1.8
The quadrivalent truth functions definable by $\neg$, $\wedge$ and $\top$ are precisely those satisfying both PERS and DUAL.

Proof: notice that the simple idea to reduce this theorem to the trivalent counterpart does not work: not every $f \in \mathsf{F}_4$ respecting PERS and DUAL is determined by its behaviour on 3^n, witness the following partially defined function f:

	1	$\frac{1}{2}$	0	2
1	1	$\frac{1}{2}$	1	2
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
0	1	$\frac{1}{2}$	1	2
2	2		2	2

where $f(2, \frac{1}{2}) = \widetilde{f(\frac{1}{2}, 2)}$ is still open. So we proceed more carefully; to this end define the auxiliary formulas $\alpha_{\vec{y},i}$ and $\beta_{\vec{y},i}$: (recall that $3 = \{0, \frac{1}{2}, 1\}$ and $\tilde{3} = \{0, 1, 2\}$)

$$\alpha_{\vec{y},i} = \begin{cases} \left.\begin{array}{ll} p_i & \text{if } y_i = 1 \\ p_i \wedge \neg p_i & \text{if } y_i = \frac{1}{2} \\ \neg p_i & \text{if } y_i = 0 \\ \top & \text{if } y_i = 2 \end{array}\right\} & \text{and } f(\vec{y}) = 1 \text{ or } \frac{1}{2} \\ \neg\top & \text{if } f(\vec{y}) = 0 \text{ or } 2 \end{cases}$$

$$\beta_{\vec{y},i} = \begin{cases} \left.\begin{array}{ll} p_i & \text{if } y_i = 1 \\ \top & \text{if } y_i = \frac{1}{2} \\ \neg p_i & \text{if } y_i = 0 \\ \top & \text{if } y_i = 2 \end{array}\right\} & \text{and } f(\vec{y}) = 1 \text{ and } \vec{y} \in 3^n \cup \tilde{3}^n \\ \neg\top & \text{otherwise} \end{cases}$$

Now let $\alpha_{\vec{y}} = \bigwedge_i \alpha_{\vec{y},i}, \beta_{\vec{y}} = \bigwedge_i \beta_{\vec{y},i}, \varphi = \bigvee_{\vec{y}} \alpha_{\vec{y}} \vee \bigvee_{\vec{y}} \beta_{\vec{y}}$. We will show that $[\![\varphi]\!]_{\vec{x}} = f(\vec{x})$ by separating the different cases: (here the default index of $[\![\cdot]\!]$ is $\vec{x}$)

[28][Vi84, p.184] uses the notation $\hat{x}$ instead of our $\tilde{x}$, and calls f *self dual* (which term we reserve for the occasion where $\tilde{f} = f$) when it is duality preserving.

[29]Or, in terms of underlying truth values: $1 \in \tilde{x} \Leftrightarrow 0 \notin x$ and $0 \in \tilde{x} \Leftrightarrow 1 \notin x$.

1. $f(\vec{x}) = 1$. Consider the following subcases:
   - if $\vec{x} \in 3^n \cup \tilde{3}^n$ then $[\![\beta_{\vec{x},i}]\!] = 1$ for all i, so $[\![\beta_{\vec{x}}]\!] = 1 \Rightarrow [\![\varphi]\!] = 1$;
   - if $\vec{x} \notin 3^n \cup \tilde{3}^n$ then $\vec{x}$ contains both $\frac{1}{2}$ and 2, say $x_j = \frac{1}{2}$ and $x_k = 2$. Now $[\![\alpha_{\vec{x},i}]\!] = \frac{1}{2}$ or 1 for all i, and $[\![\alpha_{\vec{x},j}]\!] = \frac{1}{2}$, so $[\![\alpha_{\vec{x}}]\!] = \frac{1}{2}$. Moreover (for, by DUAL, $f(\tilde{\vec{x}}) = 1$) $[\![\alpha_{\tilde{\vec{x}},k}]\!] = 2$ and $[\![\alpha_{\tilde{\vec{x}},i}]\!] = 2$ or 1 for all i, so $[\![\alpha_{\tilde{\vec{x}}}]\!] = 2$. Put together this yields $[\![\varphi]\!] = \frac{1}{2} \vee 2 \vee \cdots = 1$.

2. $f(\vec{x}) = \frac{1}{2}$. By PERS and CCLOS (cf. proposition 1.1 and footnote 32 on page 33) there is a j such that $x_j = \frac{1}{2}$ and so $[\![\alpha_{\vec{x}}]\!] = \frac{1}{2}$. We would like to show that for all $\vec{y}$ $[\![\alpha_{\vec{y}}]\!]$ and $[\![\beta_{\vec{y}}]\!]$ are either $\frac{1}{2}$ or 0. We give an indirect proof of this: assume there is a $\vec{y}$ such that either
   - $[\![\alpha_{\vec{y}}]\!] = 1$ or 2, which implies for arbitrary i: $[\![\alpha_{\vec{y},i}]\!] = 1$ or 2. So both
     (a) $f(\vec{y}) = 1$ or $\frac{1}{2}$, and
     (b) $y_i = x_i = 1$ or $y_i = x_i = 0$ or $y_i = 2$ or $x_i = 2$. In short, $\tilde{\vec{y}} \sqsubseteq \vec{x}$.

     Thus (PERS,2b) $f(\tilde{\vec{y}}) \sqsubseteq f(\vec{x}) \Rightarrow$ (DUAL,2a) $1, 2 \sqsubseteq \frac{1}{2}$, which is an absurdity.
   - or: $[\![\beta_{\vec{y}}]\!] = 1$ or $2 \Rightarrow$ for arbitrary i: $[\![\beta_{\vec{y},i}]\!] = 1$ or 2. So both
     (c) $f(\vec{y}) = 1$ & $\vec{y} \in 3^n \cup \tilde{3}^n$, and
     (d) ($y_i = 1$ & $x_i = 1, 2$) or $y_i = \frac{1}{2}$ or ($y_i = 0$ & $x_i = 0, 2$) or $y_i = 2$.

     Again we treat two cases separately:
     - $\vec{y} \in \{0, \frac{1}{2}, 1\}^n$, i.e. $y_i \neq 2$ for all i and by 2d above we obtain $\vec{y} \sqsubseteq \vec{x} \Rightarrow$ (PERS) $f(\vec{y}) \sqsubseteq f(\vec{x}) \Rightarrow$ (2c) $1 \sqsubseteq \frac{1}{2}$. Contradiction.
     - $\vec{y} \in \{0, 1, 2\}^n$, i.e. $y_i \neq \frac{1}{2}$ for all i and by 2d: $\tilde{\vec{y}} \sqsubseteq \vec{x} \Rightarrow$ (PERS,DUAL,2c) $1 \sqsubseteq \frac{1}{2}$. Contradiction.

   In all, since $[\![\alpha_{\vec{x}}]\!] = \frac{1}{2}$, $[\![\alpha_{\vec{y}}]\!] = \frac{1}{2}, 0$ and $[\![\beta_{\vec{y}}]\!] = \frac{1}{2}, 0$, we obtain $[\![\varphi]\!] = \frac{1}{2}$.

3. $f(\vec{x}) = 0$. Suppose there is a $\vec{y}$ such that $[\![\alpha_{\vec{y}}]\!] \neq 0$ or $[\![\beta_{\vec{y}}]\!] \neq 0$. Thus either
   - $[\![\beta_{\vec{y}}]\!] \neq 0$, so there is no i such that $[\![\beta_{\vec{y},i}]\!] = 0$, and moreover there is no pair j, k such that $[\![\beta_{\vec{y},j}]\!] = \frac{1}{2}$, and $[\![\beta_{\vec{y},k}]\!] = 2$. This entails that
     (a) $f(\vec{y}) = 1$ & $\vec{y} \in 3^n \cup \tilde{3}^n$, and
     (b) $y_i = 1$ & $x_i \neq 0$, or $y_i = \frac{1}{2}$, or $y_i = 0$ & $x_i \neq 1$, or $y_i = 2$.
     (c) either for all i such that $y_i \in \{0, 1\} : x_i \neq \frac{1}{2}$ or all such i: $x_i \neq 2$.

     Then distinguish:
     - $\vec{y} \in \{0, \frac{1}{2}, 1\}^n$, i.e. $y_i \neq 2$ for all i and by 3b and 3c we obtain $\vec{y} \sqsubseteq \vec{x}$ or $\vec{y} \sqsubseteq \tilde{\vec{x}} \Rightarrow$ (PERS,DUAL,3a) $1 \sqsubseteq 0$. Contradiction.
     - $\vec{y} \in \{0, 1, 2\}^n$, i.e. $y_i \neq \frac{1}{2}$ for all i and by a similar argument: $\vec{x} \sqsubseteq \vec{y}$ or $\tilde{\vec{x}} \sqsubseteq \vec{y} \Rightarrow 0 \sqsubseteq 1$. Contradiction.
   - $[\![\beta_{\vec{y}}]\!] = 0$ and $[\![\alpha_{\vec{y}}]\!] \neq 0$. This implies that each $[\![\alpha_{\vec{y},i}]\!]$ is non-zero and does not take the values $\frac{1}{2}$ and 2 for different i. So $f(\vec{y}) = \frac{1}{2}$ or $1 \Rightarrow f(\vec{x}) \not\sqsubseteq f(\vec{y}) \Rightarrow$ (PERS) $\vec{x} \not\sqsubseteq \vec{y} \Rightarrow$ for some $k : x_k \not\sqsubseteq y_k$, and therefore $x_k \neq \frac{1}{2}$, which leaves us with the following possiblities:
     - $x_k = 0 \Rightarrow y_k = \frac{1}{2}, 1$, but then $[\![\alpha_{\vec{y},k}]\!] = 0$. Contradiction

- $x_k = 1 \Rightarrow y_k = \frac{1}{2}, 0$ and $[\![\alpha_{\vec{y},k}]\!] = 0$. Contradiction
- $x_k = 2 \Rightarrow y_k \neq 2$, $[\![\alpha_{\vec{y},k}]\!] = 2$. So for all i $[\![\alpha_{\vec{y},i}]\!] = 1$ or 2, i.e. ($y_i = 1$ & $x_i = 1, 2$) or ($y_i = \frac{1}{2}$ & $x_i = 2$) or ($y_i = 0$ & $x_i = 0, 2$) or $y_i = 2$. So $\tilde{\vec{x}} \sqsubseteq \vec{y}$ $\Rightarrow 0 \sqsubseteq f(\vec{y})$. Contradiction.

The joint effect is that in this case $[\![\varphi]\!] = 0$.

4. $f(\vec{x}) = 2$ Since $f \in$ DUAL we know that $f(\tilde{\vec{x}}) = \frac{1}{2}$. By case 2 we obtain that $[\![\varphi]\!]_{\tilde{\vec{x}}} = \frac{1}{2}$. Also, since $\top$, $\neg$ and $\wedge$ are duality preserving, it follows by induction that $[\![\varphi]\!]_{\tilde{\vec{x}}} = \widetilde{[\![\varphi]\!]_{\vec{x}}}$. Therefore $[\![\varphi]\!]_{\vec{x}} = \widetilde{[\![\varphi]\!]_{\tilde{\vec{x}}}} = \tilde{\frac{1}{2}} = 2$. ∎

This immediately suggests a closely related result characterizing the basic connectives $\neg$ and $\wedge$, reaching our temporary end goal, the four-valued counterpart of theorem 1.5.

Theorem 1.9
The quadrivalent truth functions definable by $\neg$ and $\wedge$ are precisely those satisfying the combination of conditions PERS, DUAL and FREE.

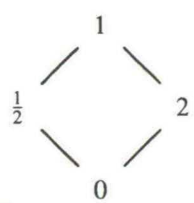
Proof: referring to the previous proof, turn the full definitions of $\alpha_{\vec{y},i}$ and $\beta_{\vec{y},i}$ into partial ones by dropping the cases $\top$ and $\neg\top$. The formula φ cannot be void, since $f(\vec{\frac{1}{2}}) = \frac{1}{2}$, so α contains at least the disjunct $\bigwedge\!\!\!\bigwedge_i (p_i \wedge \neg p_i)$. ∎

This is the end of the main road through this section; we do want to return shortly to some sideways which also lead to nice spots.

Given the crucial effect of DUAL we may wonder which connectives characterize this condition alone. Consider the new negation[30] symbolized by ∂:

	1	$\frac{1}{2}$	0	2
∂	0	2	1	$\frac{1}{2}$

It may be argued that this *dual negation* ∂, despite its typical 4-valued appearance, gives a classical ring to the system. To illustrate this, note that $\varphi \vee \partial\varphi$ is always true (i.e. either 1 or 2). In this respect ∂ is very much like $\sim$, but there are differences too, as will become clear soon. In fact much of this can already be understood once we notice the effect of ∂ on the logical lattice $\langle 4, \wedge, \vee \rangle$: it is turned into a genuine *Boolean Algebra*.



[30]Incidentally, this operator suggested itself by a print error in [Be77, p.13], where the truth table for $\neg$ coincides with the one given for ∂ here, obviously conflicting with the intended persistence. Yet the operation corresponding to ∂ is known as *involution* in relevance logic, cf. [Du86].

In a sense ∂ is stronger than $\sim$ since together with $\neg$, $\wedge$ and $\star$ it already defines F_4.[31] Without $\star$ these operators capture DUAL:

Theorem 1.10
Every duality preserving 4-valued truth function is definable by $\{\neg, \wedge, \partial\}$.

Proof: First some heuristics. Consider the two-place function in DUAL characterized by the matrix

$$\begin{array}{c|cccc} & 1 & \frac{1}{2} & 0 & 2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 \end{array}$$

Though the unary function which assigns value 1 for arguments $\frac{1}{2}$ and 2 and has as many zero-values as possible, i.e. $\begin{array}{|cccc} 1 & \frac{1}{2} & 0 & 2 \\ \hline 0 & 1 & 0 & 1 \end{array}$, can be characterized by the given connectives, this does not help: the (conjunctive) product of two such unary functions will result in the wrong matrix

$$\begin{array}{c|cccc} & 1 & \frac{1}{2} & 0 & 2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 \end{array}$$

Therefore, we employ a different strategy: 'split' the initial matrix into the pair

$$\begin{array}{c|cccc} & 1 & \frac{1}{2} & 0 & 2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 \end{array} \quad \& \quad \begin{array}{c|cccc} & 1 & \frac{1}{2} & 0 & 2 \\ \hline 1 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & \frac{1}{2} & 0 & 0 \end{array},$$

and notice that the initial matrix is the sum of the last two. Moreover, the latter matrices can be considered the two products of the matrices corresponding to the functions

$$\begin{array}{|cccc} 1 & \frac{1}{2} & 0 & 2 \\ \hline 0 & \frac{1}{2} & 0 & 2 \end{array} \qquad \begin{array}{|cccc} 1 & \frac{1}{2} & 0 & 2 \\ \hline 0 & 2 & 0 & \frac{1}{2} \end{array}$$

Now consider a number of expressions that are 'most specific':

p	$p \wedge \neg\partial p$	$p \wedge \neg p$	$\neg p \wedge \partial p$	$\partial p \wedge \neg\partial p$	$p \wedge \partial p$
1	1	0	0	0	0
$\frac{1}{2}$	0	$\frac{1}{2}$	0	2	0
0	0	0	1	0	0
2	0	2	0	$\frac{1}{2}$	0

[31] This is a consequence of theorem 1.6 and the reductions: $\natural = \partial\star$ and $\sim\varphi = (\neg\varphi \wedge \star) \vee (\partial\varphi \wedge \natural)$.

By means of these expressions we define, for any mapping f that is duality preserving, a characteristic formula φ that maximally exploits DUAL:

$$\varphi_{\vec{y},i} = \begin{cases} \left.\begin{array}{ll} p_i \wedge \neg\partial p_i & \text{if } y_i = 1 \\ p_i \wedge \neg p_i & \text{if } y_i = \frac{1}{2} \\ \neg p_i \wedge \partial p_i & \text{if } y_i = 0 \\ \partial p_i \wedge \neg\partial p_i & \text{if } y_i = 2 \end{array}\right\} & \text{and } f(\vec{y}) = \frac{1}{2} \text{ or } 1 \\ p_i \wedge \partial p_i & \text{if } f(\vec{y}) = 0 \text{ or } 2 \end{cases}$$

$$\varphi = \bigvee_{\vec{y}} \bigwedge_{i=1}^{i=n} \varphi_{\vec{y},i}$$

To show that $[\![\varphi]\!]_{\vec{x}} = f(\vec{x})$ we start with a useful observation (where $\varphi_{\vec{y}} = \bigwedge_{i=1}^{i=n} \varphi_{\vec{y},i}$):

observation If $f(\vec{y}) = \frac{1}{2}$ or 1, then $[\![\varphi_{\vec{y}}]\!] = \frac{1}{2}$ or 1 when $\vec{y} = \vec{x}$, and $[\![\varphi_{\vec{y}}]\!] = 1$ or 2 when $\vec{y} = \tilde{\vec{x}}$; in all other cases $[\![\varphi_{\vec{y}}]\!] = 0$.

Once again we treat the different values of $f(\vec{x})$ separately:

1. if $f(\vec{x}) = 1$ there are two subcases:
   (a) $\vec{x} \in 2^n \Rightarrow [\![\varphi_{\vec{x},i}]\!] = 1$ for all $i \Rightarrow [\![\varphi_{\vec{x}}]\!] = 1 \Rightarrow [\![\varphi]\!] = 1$;
   (b) if $\vec{x} \notin 2^n$ then for some j: $x_j \notin 2$, so $[\![\varphi_{\vec{x},j}]\!] = \frac{1}{2}$ and $[\![\varphi_{\tilde{\vec{x}},j}]\!] = 2$, whereas for other i (cf. the observation) $[\![\varphi_{\vec{x},i}]\!] = \frac{1}{2}$ or 1, and $[\![\varphi_{\tilde{\vec{x}},i}]\!] = 2$ or 1, and therefore $[\![\varphi_{\vec{x}}]\!] = \frac{1}{2}$ and $[\![\varphi_{\tilde{\vec{x}}}]\!] = 2$. So in this case too $[\![\varphi]\!] = \frac{1}{2} \vee 2 \vee \cdots = 1$.
2. $f(\vec{x}) = \frac{1}{2}$. By proposition 1.1, $f \in$ CCLOS, thus $\vec{x} \notin 2^n$. Similar to case 1b we know that $[\![\varphi_{\vec{x}}]\!] = \frac{1}{2}$ and moreover for any other $\vec{y}$: $[\![\varphi_{\vec{y}}]\!] = 0, \frac{1}{2}$ for suppose, by contrast, that there is a $\vec{y} \neq \vec{x}$ such that $[\![\varphi_{\vec{y}}]\!] = 1$ or 2, then, witness the observation above, $f(\vec{y}) = \frac{1}{2}$ or 1 and $\vec{y} = \tilde{\vec{x}}$, and therefore (DUAL) $f(\vec{y}) = 2$. Contradiction. So $[\![\varphi]\!] = \frac{1}{2}$.
3. $f(\vec{x}) = 0$. Suppose there is a $\vec{y}$ such that $[\![\varphi_{\vec{y}}]\!] \neq 0$. Thus the observation tells us that $f(\vec{y}) = \frac{1}{2}$ or 1, and $\vec{y} = \vec{x}$ or $\vec{y} = \tilde{\vec{x}}$, and so (by DUAL) $f(\vec{y}) = 0$. Contradiction.
4. $f(\vec{x}) = 2$ Since $f \in$ DUAL, $f(\tilde{\vec{x}}) = \frac{1}{2}$. By case 2 we obtain that $[\![\varphi]\!]_{\tilde{\vec{x}}} = \frac{1}{2}$. Also, since ∂, $\neg$ and $\wedge$ are duality preserving, it follows by induction that $[\![\varphi]\!]_{\vec{x}} = \widetilde{[\![\varphi]\!]_{\tilde{\vec{x}}}} = \tilde{\frac{1}{2}} = 2$.

■

We may also study the effect of CCLOS in isolation. Again incorporation of dual negation is helpful. One way of transferring Langholm's and van Benthem's earlier result to the 4-valued case thus becomes:

Theorem 1.11
Every classically closed quadrivalent truth function is definable by $\{\neg, \wedge, \sim, \partial\}$.

Proof: first notice that the listed operators respect CCLOS. We can now proceed as in theorem 1.6, i.e. for some given closed function f produce subformulas characteristic for $\langle \vec{x}, f(\vec{x}) \rangle$. Apart from the auxiliary formulas χ_{1000}, χ_{0100}, χ_{0010} and χ_{0001} we introduce:

p	$\chi_{0\frac{1}{2}00}(p) =$ $p \wedge \sim p$	$\chi_{0200}(p) =$ $\sim p \wedge \neg\partial p$	$\chi_{000\frac{1}{2}}(p) =$ $\partial p \wedge \neg\sim p$	$\chi_{0002}(p) =$ $\neg p \wedge \neg\sim p$	$\chi_{0000}(p) =$ $\sim p \wedge \neg\sim p$
1	0	0	0	0	0
$\frac{1}{2}$	$\frac{1}{2}$	2	0	0	0
0	0	0	0	0	0
2	0	0	$\frac{1}{2}$	2	0

The conjuncts $\varphi_{\vec{x},i}$ are defined by distinguishing the following cases:

- if $f(\vec{x}) = 1$ then $\varphi_{\vec{x},i} = \begin{cases} \chi_{1000}(p_i) & \text{if } x_i = 1 \\ \chi_{0100}(p_i) & \text{if } x_i = \frac{1}{2} \\ \chi_{0010}(p_i) & \text{if } x_i = 0 \\ \chi_{0001}(p_i) & \text{if } x_i = 2 \end{cases}$

- if $f(\vec{x}) = \frac{1}{2}$ then $\varphi_{\vec{x},i} = \begin{cases} \chi_{1000}(p_i) & \text{if } x_i = 1 \\ \chi_{0\frac{1}{2}00}(p_i) & \text{if } x_i = \frac{1}{2} \\ \chi_{0010}(p_i) & \text{if } x_i = 0 \\ \chi_{000\frac{1}{2}}(p_i) & \text{if } x_i = 2 \end{cases}$

- if $f(\vec{x}) = 0$ then $\varphi_{\vec{x},i} = \chi_{0000}(p_i)$

- if $f(\vec{x}) = 2$ then $\varphi_{\vec{x},i} = \begin{cases} \chi_{1000}(p_i) & \text{if } x_i = 1 \\ \chi_{0200}(p_i) & \text{if } x_i = \frac{1}{2} \\ \chi_{0010}(p_i) & \text{if } x_i = 0 \\ \chi_{0002}(p_i) & \text{if } x_i = 2 \end{cases}$

Now the formula corresponding to f is

$$\varphi = \bigvee_{\vec{x}} \bigwedge_{i=1}^{i=n} \varphi_{\vec{x},i}.$$

In order to check that $[\![\varphi]\!]_{\vec{x}} = f(\vec{x})$, we first calculate $[\![\bigwedge_{i=1}^{i=n} \varphi_{\vec{y},i}]\!]_{\vec{x}}$:

- if $\vec{y} \neq \vec{x}$, then there is an i such that $y_i \neq x_i \Rightarrow [\![\varphi_{\vec{y},i}]\!] = 0$ (for example, if $y_i = 2 \neq x_i$ then $[\![\varphi_{\vec{y},i}]\!] = [\![\chi_{000z}(p_i)]\!] = 0$) and so $[\![\bigwedge_i \varphi_{\vec{y},i}]\!] = 0$;
- if $\vec{y} = \vec{x}$ then
  - if $f(\vec{x}) = 1$ (or 0) then by the definition of $\varphi_{\vec{x},i}$: $[\![\varphi_{\vec{x},i}]\!] = 1$ (0) for all i, and so $[\![\bigwedge_i \varphi_{\vec{x},i}]\!] = 1$ (or 0, respectively);
  - if $f(\vec{x}) = \frac{1}{2}$ then by CCLOS there must an i such that $x_i = \frac{1}{2}$ or 2, and furthermore $[\![\varphi_{\vec{x},i}]\!] = \frac{1}{2}$ or 1 for other i, and so $[\![\bigwedge_i \varphi_{\vec{x},i}]\!] = \frac{1}{2}$;
  - if $f(\vec{x}) = 2$ then by a similar argument $[\![\bigwedge_i \varphi_{\vec{x},i}]\!] = 2$.

In all, $[\![\varphi]\!] = [\![\bigwedge \varphi_{\vec{x},i}]\!]_{\vec{x}} = f(\vec{x})$. ∎

This is in fact only one of the counterparts of the 3-valued CCLOS–theorem: we may opt for the other perspective on definability and start with the set of connectives $\{\neg, \wedge, \sim\}$ and ask for the characteristic condition. The preceding theorem indicates that we need a stronger condition to capture these connectives. Notice that there are several ways to modify or extend CCLOS, but one that is particularly appropriate for the quadrivalent case is what we call *general closure*: (recall that $3 = \{0, \frac{1}{2}, 1\}$ and $\tilde{3} = \{0, 1, 2\}$)

GCLOS f is *generally closed* iff $f[3^n] \subseteq 3$ and $f[\tilde{3}^n] \subseteq \tilde{3}$.

GCLOS obviously implies CCLOS. Since ∂ is CCLOS but not in GCLOS, we know that this condition is really stronger than CCLOS.[32] Because $\neg$, $\wedge$ and $\sim$ are in CCLOS, we also obtain that ∂ is not redundant in the previous theorem. The latter connectives are in fact functionally complete for GCLOS, so we may argue that this is the correct counterpart of CCLOS in the 4-valued case.

Theorem 1.12
Every generally closed quadrivalent truth function is definable by $\{\neg, \wedge, \sim\}$.

Proof: in fact it suffices to reinspect the proof of theorem 1.11. The only characteristic formulas using ∂ were χ_{0200} and $\chi_{000\frac{1}{2}}$. The latter formulas are not in GCLOS and in defining $\varphi_{\bar{x},i}$ should be replaced by χ_{0100} and χ_{0001}, respectively. The rest of the proof is analogous to the previous one. ∎

Although we succeeded in giving definability results which formed the main road through the jungle of four-valued semantics, we do not have similar results for every possible combination of conditions.[33] Especially the condition *freedom* is still quite indeterminate in this respect, but here we should ask ourselves whether we really want to be dragged into such an exhaustive experience: without *persistence*, freedom is hardly interesting.

Apart from mere definability, we may also want to compare the conditions for their relative strength. Again this problem can be approached in a way that is reminiscent of Generalized Quantifier Theory: count the number of functions they single out. This will be the subject of the next section, which can be skipped without loss of continuity.

1.4 The strength of conditions: counting results

Probably one of the best ways to compare the strength of conditions is to measure the size of the class of functions of fixed arity fulfilling the requirements. For example, there are $4^{16} \approx 4.3 \cdot 10^9$ two-place quadrivalent functions, of which $4^{15} \approx 1.1 \cdot 10^9$ are free, $4^{14} \approx 2.7 \cdot 10^8$ are classically closed, $4^8 = 65\,536$ preserve dualization and 28 224 are persistent. Also the numerical effect of combinations of conditions can be interesting, for such conditions can act independently, reinforce or weaken each other. For example, out of the 4-valued two-place functions $4^{13} \approx 6.7 \cdot 10^7$ are free and

[32] GCLOS and CCLOS are, however, equivalent modulo PERS, see [Th90c, prop.5.2].

[33] For example, which connectives characterize PERS + CCLOS ?

classically closed, yet no more than 168 respect PERS + DUAL, almost all of which (166) are also free.

Instead of calculating through every single case, we would like to have general formulas for combinations of conditions, or at least provide upper and lower bounds. For any combination of conditions C, the number of n-ary k-valued functions fulfilling C will be denoted by $|\mathrm{C}_k^{(n)}|$. It is obvious that $|\mathrm{F}_k^{(n)}| = k^{k^n}$ and $|\mathrm{FREE}_k^{(n)}| = k^{k^n-1}$ ($k = 3, 4$). Here are results about closure properties:

Proposition 1.2

1. $|\mathrm{CCLOS}_k^{(n)}| = 2^{2^n} \cdot k^{(k^n-2^n)}$;

2. $|\mathrm{GCLOS}_4^{(n)}| = 4^{(4^n-2\cdot 3^n+3\cdot 2^{n-1})} \cdot 9^{(3^n-2^n)}$.

Proof: straightforward from the definitions, for example, the calculation for GCLOS rests on a partition of the argument space 4^n into groups which can get the same values: for arguments $\vec{x}$ in 2^n all and only values in $2 = \{0, 1\}$ qualify; for arguments in $3^n - 2^n$ all and only values in 3; for $\tilde{3}^n - 2^n$ in $\tilde{3}$ and finally for $4^n - (3^n \cup \tilde{3}^n)$ all members of 4 qualify. Counting numbers of these independent choices produces the total number: $2^{2^n} \cdot 3^{3^n-2^n} \cdot 3^{3^n-2^n} \cdot 4^{4^n-2\cdot 3^n+2^n} = 4^{(4^n-2\cdot 3^n+3\cdot 2^{n-1})} \cdot 9^{(3^n-2^n)}$. ∎

The addition of *freedom* to either CCLOS or GCLOS reduces the relevant exponent in the predictable way.

More interesting and less obvious is the behaviour of PERS and DUAL. For $k = 3$ [B186] gives some numbers: $|\mathrm{PERS}_3^{(0)}| = 3$, $|\mathrm{PERS}_3^{(1)}| = 11$, $|\mathrm{PERS}_3^{(2)}| = 197$, and (calculated by A. W. Roscoe) $|\mathrm{PERS}_3^{(3)}| = 129\,615$, $|\mathrm{PERS}_3^{(4)}| = 430\,904\,428\,717 \approx 4.3 \cdot 10^{11}$. So far, nobody has been able to give an explicit formula predicting this sequence. We do not have a solution for this problem, but we do have a formula relating the case for $k = 4$ to a well-known unsolved mathematical puzzle. Let Δ_n be the number of monotonically increasing generalized quantifiers on the domain $\{1, \ldots, n\}$, the 'Dedekind number' of n.[34] Then the problem for $|\mathrm{PERS}_4^{(n)}|$ can be reduced to Dedekind's problem witness the next result, where PERS is also compared and related to DUAL:

Proposition 1.3

1. $|\mathrm{PERS}_4^{(n)}| = (\Delta_{2n})^2$

2. $|\mathrm{DUAL}_4^{(n)}| = 2^{4^n}$

3. $|\mathrm{PERS} \cap \mathrm{DUAL}_4^{(n)}| = \Delta_{2n}$

4. $|\mathrm{DUAL} \cap \mathrm{FREE}_4^{(n)}| = 2^{4^n-2}$;

[34] A generalized quantifiers Q on $\{1, \ldots, n\}$ is a subset of $\wp\{1, \ldots, n\}$. Q is monotonically increasing iff $A \subseteq B$ & $A \in Q \Rightarrow B \in Q$. Cf. [Th85] for background and explanation.

5. $|\text{PERS} \cap \text{DUAL} \cap \text{FREE}_4^{(n)}| = \Delta_{2n} - 2$;

Proof:

1. The lattice $\langle 4^n, \sqsubseteq\rangle$ is isomorphic to the lattice $\langle \wp(\{0,1\} \times \{1,\ldots,n\}), \subseteq\rangle$ since each $\vec{x} \in \{0, \frac{1}{2}, 1, 2\}$ corresponds to a set

$$S_{\vec{x}} = \{\langle a, i\rangle \mid a = 0, 1 \;\&\; 1 \leq i \leq n \;\&\; a \sqsubseteq x_i\}.$$

It is easy to see that this correspondence is one-to-one (for example, that it is onto follows from building up co-ordinates x_i by consideration of membership of the underlying truth values, combined with i) and preserves the partial order. Furthermore every persistent function $f : 4^n \longrightarrow 4$ can be encoded in two *independent* monotonic subsets of $\wp(\{0,1\} \times \{1,\ldots,n\})$, or rather their characteristic functions f_0 and f_1, which are defined by $f_i(S_{\vec{x}}) = 1 \Leftrightarrow i \sqsubseteq f(\vec{x})$.

2. $|\text{DUAL}_4^{(n)}| = 2^{2^n} \cdot 4^{(4^n - 2^n)/2} = 2^{4^n}$.

3. Using the proof of 1, this is a straightforward consequence of the fact that f is now determined by just one of f_0 and f_1, witness the following equivalences:

$$f_0(S_{\vec{x}}) = 1 \Leftrightarrow 0 \sqsubseteq f(\vec{x}) \Leftrightarrow 1 \not\sqsubseteq \widetilde{f(\vec{x})} \Leftrightarrow 1 \not\sqsubseteq f(\tilde{\vec{x}}) \Leftrightarrow f_1(S_{\tilde{\vec{x}}}) = 0.$$

4. $|\text{DUAL} \cap \text{FREE}_4^{(n)}| = 2^{2^n} \cdot 4^{(4^n - 2^n - 2)/2} = 2^{4^n - 2}$.

5. Notice that now only two functions are eliminated when adding FREE to PERS and DUAL. First focus on PERS + DUAL. By the implied GCLOS we have $f(\vec{\tfrac{1}{2}}) \in 3$; if $f(\vec{\tfrac{1}{2}}) = 1$ we also have (by DUAL) $f(\vec{2}) = 1$ and so by persistence $f(\vec{x}) = 1$ for all $\vec{x}$. And, similarly if $f(\vec{\tfrac{1}{2}}) = 0$, $f(\vec{x})$ is constantly 0. ∎

The result immediately provides some values: $|\text{PERS}_4^{(0)}| = 4$, $|\text{PERS}_4^{(1)}| = 36$, $|\text{PERS}_4^{(2)}| = 168^2 = 28\,224$ and $|\text{PERS}_4^{(3)}| = 7\,828\,354^2 \approx 6.1 \cdot 10^{13}$ (the last may seem much, but is still less than for example $|\text{DUAL}_4^{(3)}| = 2^{64} \approx 1.8 \cdot 10^{19}$). Yet a general expression in terms of standard arithmetical operations is not known for Δ_n. Although this relation to a longstanding open problem may be interesting, for the purpose of this section it is more important to estimate the number of quadrivalent persistent functions than to assess their precise amount. Using well-known results about the Dedekind number, we arrive at the following approximations:[35]

Proposition 1.4

1. $4^{\binom{2n}{n}} \leq |\text{PERS}_4^{(n)}| \leq 9^{\binom{2n}{n}}$

[35] A further (rough) approximation is $\binom{2n}{n} \approx \frac{4^n}{\sqrt{\pi n}}$. The absolute deviation of this approximation increases, but the relative one decreases for large n.

2. $2^{\binom{2n}{n}} \leq |\mathrm{PERS} \cap \mathrm{DUAL}_4^{(n)}| \leq 3^{\binom{2n}{n}}$

So for the 4-valued case the strength of the single conditions is ordered as follows:

$$\mathrm{PERS} \gg \mathrm{DUAL} \gg \mathrm{GCLOS} > \mathrm{CCLOS} > \mathrm{FREE} > \textit{no condition}$$

To recapitulate, we have shown the precise relation between some counting tasks related to persistence on the one hand, and an open combinatorical problem on the other. Moreover, we were able to assess the relative strength of conditions used in isolation or in combination. Finally, note that there is no point in comparing the order of the numbers $|\mathrm{C}_k^{(n)}|$: they all are superexponential. But of course this is what makes definability worth while: the enormous expressive potential of small sets of logical constants.

1.5 Conclusion

In this chapter a number of conditions on (3- and 4-valued) truth functions have been discussed in relation to certain connectives. Several combinations of conditions have been shown to correspond to groups of operators. This, of course, raises the question of the success of this enterprise. Moreover, since there is no way in which we could deal with *every* possible condition or connective[36], we may wonder whether we have really treated the most evident and interesting cases.

To start with the latter question, we believe that indeed we have: the proposed conditions[37] seem quite natural and moreover characterize sets of independently motivated connectives.

As regards the former question of the success of the definability task, notice that, although the picture is by no means complete, we are quite satisfied, since the most evident cases have been covered.

Moreover, our route through the field of possible results was motivated by the attempt to characterize the classical connectives within partial semantics; this was achieved by successively adding conditions, first for the three-valued models:

no condition:	$\neg, \wedge, (\top), \star, \sim$
persistence:	$\neg, \wedge, \top, \star$
persistence + classical closure:	$\neg, \wedge, \top$
persistence + classical closure + freedom:	$\neg, \wedge$

This is mirrored in the four-valued case by a similar sequence:

no condition:	$\neg, \wedge, (\top), \star, \sharp, \sim$
persistence:	$\neg, \wedge, (\top), , \star, \sharp$
persistence + duality preservation:	$\neg, \wedge, \top$
persistence + duality preservation + freedom:	$\neg, \wedge$

[36] Unless in such general terms as in proposition 1.7 of the appendix.

[37] In the appendix we discuss some other conditions suggested in the literature.

Apart from this main road, there were some miscellaneous results. First, for the three-valued truth functions:

classical closure: $\neg, \wedge, \sim$
persistence + freedom: $\neg, \wedge, \star$

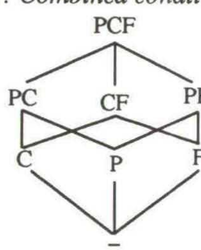
Second, for four-valued truth functions:

classical closure: $\neg, \wedge, \sim, \partial$
general closure: $\neg, \wedge, \sim$
duality preservation: $\neg, \wedge, \partial$

To see to what extent we have covered the possibilities, we may again employ the twofold perspective mentioned in the introduction.

Reasoning from conditions to connectives, we note that out of the 8 possible combinations for the trivalent conditions PERS, CCLOS and FREE, only 6 have been shown to be definable by sets of connectives (leaving open the combinations FREE and CCLOS+FREE). We display the order of the combined conditions for F_3 in figure 1.1 (a condition is abbreviated as its first letter).

Figure 1.1: *Combined conditions for* F_3

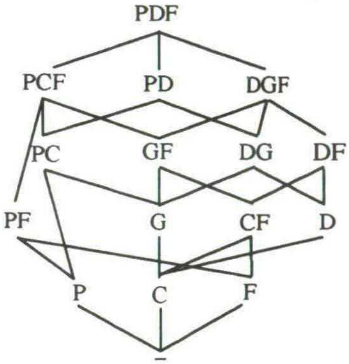


However, if we take the classical connectives $\neg$ and $\wedge$ to be a fixed kernel (as we have done before), the 8 possible combinations with $\star$, $\top$ and $\sim$ *are* describable by PERS, CCLOS and FREE. How can this be? The simple reason is that $\top$ can be defined by $\sim$, so there are two redundancies in the general scheme. Hence, figure 1.2 shows full success of the characterization task for the connectives encountered in the three-valued case. It also implies that the open problems with regard to figure 1.1 can only be resolved by introducing new connectives.

For the quadrivalent case definability is more complicated. First notice that PERS, DUAL, GCLOS, CCLOS and FREE do not present 32 extensionally different subsets: since GCLOS implies CCLOS, and DUAL implies CCLOS, certain combinations can be identified, and only 16 survive (figure 1.3). So far only 7 out of these 16 combinations have been covered by definability results. Although we have some ideas how to fill these gaps (involving new conditions and new connectives), this will not be pursued here.

Figure 1.2: *Characterization of connectives in* F_3

Figure 1.3: *Combined conditions for* F_4



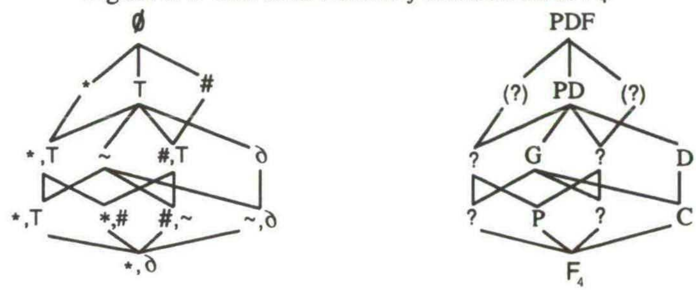
Starting with $\neg$ and $\wedge$, addition of $\star$, $\sharp$, $\top$, $\sim$ and ∂ improves the situation: after removing redundancies, 7 out of 13 different systems are definable by the stated conditions, see figure 1.4.

Despite the open spots in this diagram, we conclude that the most relevant results have been added to the knowledge of the subject. Definability results provide a deeper understanding of the expressiveness of the language, and guides the choice of the connectives in the subsequent chapters. Moreover, this kind of definability is useful for answering questions like: Is connective c_0 definable by $c_1, \ldots, c_n$? We do not have to go through intricate syntactic arguments, especially in the case of a negative answer. For example, it now immediately follows that $\sim$ is not definable by $\neg$, $\wedge$ and ∂, since the latter all preserve duality whereas the former does not.

Appendix: which conditions are definable?

There were some issues omitted from the main text which do not belong to the kernel of this thesis, but may nevertheless be of interest to some readers. The first subsection deals with the precise definition of what constitutes a proper condition, with the kind of subtlety familiar from basic recursion theory. This may be of interest

Figure 1.4: *Characterization of connectives in* F_4



to the mathematically minded. After treating one alternative in some more detail, we recapitulate some other proposed conditions, to serve the more historically minded.

some general definitions

The search for proper conditions is constrained by the fact that, in order to be definable, classes of functions need to reflect some basic properties of the syntax of propositional logic.

- Well-formed formulas (WFFs) can reoccur in more complex formulas, i.e. the notion of WFF is recursive.
- The propositional atoms may occur within a formula in any order, may or may not be used, and may reoccur.

These simple facts correspond to semantic properties, and these constraints will delimit the set of proper classes of truth functions.

Definition 1.4 (projection)
For any $i = 1, \ldots, n$ *the function* $\pi_i : k^n \longrightarrow k$ *is a* **projection** *iff for all* $\vec{x} \in k^n$: $\pi_i(\vec{x}) = x_i$.

Definition 1.5 (generalized composition)
If $f \in k^{k^m}$ *and* $g_1, \ldots, g_m \in k^{k^n}$ *then* $f \circ g_1 g_2 \cdots g_m \in k^{k^n}$ *is the* **generalized composition** *of* f *and* $g_1, \ldots, g_m$, *which is defined by* $f \circ g_1 \ldots g_m(\vec{x}) = f(g_1(\vec{x}), \ldots, g_m(\vec{x}))$ *for all* $\vec{x} \in k^n$.

Definition 1.6 (closed class)
$\mathsf{C} \subseteq \mathsf{F}_k$ *is a* **closed class** *iff for each* $n > 0$ *it contains the projections* $\pi_1, \ldots, \pi_n \in k^{k^n}$ *and is closed under generalized composition.*

The extended notion of composition may seem incapable of describing arbitrary composition as is needed for closed classes of truth functions. For example, the complex function given by $\langle x, y, z\rangle \mapsto f(g(y,z), h(x), i(z,x))$ combines functions of arity 3,2,1 and 2, respectively. However, we can describe this mapping by $f \circ (g \circ \pi_2\pi_3)(h \circ \pi_1)(i \circ \pi_3\pi_1)$ which belongs to k^{k^3} by recursive use of the definition of composition. This function could correspond to a logical formula such as $(q \wedge r) \vee \neg p \vee (r \rightarrow p)$; this correspondence becomes especially clear when we use the 'Cambridge Polish' notation with many–place disjunctions: $(\vee(\wedge qr)(\neg p)(\rightarrow rp))$.

Here are some simple, yet useful facts about closed classes:

Proposition 1.5

- *Combination of conditions corresponding to closed classes corresponds to the intersection of these classes.*
- *The intersection of closed classes is a closed class.*
- *Every closed class contains functions of arbitrary arity $n > 0$, viz. the n projections.*

From the above definitions it also follows that there are two trivially closed classes in the case of k truth values:

1. the whole function space F_k;
2. the set of all projections; for it is easy to see that a composition of projections is a projection again.

These two trivial classes are the simplest illustrations; they are the top and bottom elements in a lattice of closed classes, ordered by ordinary set inclusion. The interesting cases are somewhere in between. Also, we can now easily give examples of classes which are not closed:

Proposition 1.6

- *A class of functions of fixed arity is not closed.*
- *The complement of a closed class is not closed.* [38]

The vague ideas mentioned in the beginning of this appendix can now be formalized by the following proposition:

Proposition 1.7 $\mathsf{C} \subseteq \mathsf{F}_k$ *is definable by some C iff* C *is a closed class; moreover,* C *is definable by a* **finite** *C iff* C *is the closure under compositions of the projections and a* **finite** *set of other functions.*

[38] Nor does union preserve closure of classes: $\neg \sim p \wedge \star$ is neither classically closed nor free though its defining operators are either way.

Of course this apparatus may be used to justify a condition, but also to reject a condition, if it does not meet the standards of definability. Below we discuss the possible condition of distributivity, providing a non-trivial illustration of the notions introduced.

distributivity

Distributivity is an interesting condition since it motivates the strong Kleene extensions of the binary truth tables for the classical connectives.[39]

First we reconsider the operations $\sqcap$ (meet) and $\sqcup$ (join) which are the greatest lower bound and the lowest upper bound in the approximation lattice. These operations correspond to the intersection and union of underlying truth value sets.

Meet and join can be generalized in the usual way to vectors (by pointwise definition) and sets (notation $\bigsqcap$ and $\bigsqcap$). Then *distributivity* can get its formal description:

DISTR f is distributes over $\bigsqcup$ iff $f(\bigsqcup X) = \bigsqcup f[X_1 \times \cdots \times X_n]$,

where $X \subseteq 4^n$ and $X_i = \pi_i[X]$. The definition of distributivity with respect to $\bigsqcap$ is completely analogous (and may be applied to 3^n as well).

One simple case is that in which X contains only two members which coincide except for one co-ordinate, then for a binary function f:

$$f(x, y \sqcup z) = f(x, y) \sqcup f(x, z).$$

In fact distributivity can be defined from this restricted form by iteration. Another easy case is when f is a unary function; then the criterion simplifies to $f(\bigsqcup X) = \bigsqcup f[X]$. In fact one may wonder whether this transparant equation can replace the one in the above definition, for n–ary functions. But there are counterexamples to this idea, for example, let f be the interpretation of $\wedge$, then

$$f(\langle 1,0\rangle \sqcup \langle 0,1\rangle) = f(2,2) = 2 \neq f(1,0) \sqcup f(0,1) = 0 \sqcup 0 = 0.$$

Therefore the simplified equation is in general incorrect.[40]

Now what are the connectives defining distributivity, *if any*? Since the usual connectives are distributive with respect to $\sqcap$ (and $\sqcup$ in the 4-valued case), the expressiveness result suggested by this is:

Conjecture 1.1
The distributive functions can be defined using only $\neg$ and $\wedge$.

However, this conjecture is wrong — in fact it fails in both directions:

---

[39] See [Th90c] for a lengthy demonstration of this.

[40] Yet, it is this equation that is proposed by [Be77]. Belnap requires X to be *directed*, i.e. confluent in the right direction; for $\sqcup$ the requirement is that for any $x, y \in X$ there is a $z \in X$ such that $x \sqsubseteq z$ and $y \sqsubseteq z$. Referring to Scott, he calls the property of distributivity for directed sets *continuity*, which is, for finite X, equivalent to *persistence*.

- the distributive function $f : x \mapsto 1$ is not definable in terms of standard connectives and propositional variables (for motivation see theorems 1.5 and 1.9).
- the function f with $f(1) = f(0) = 0$ and $f(\frac{1}{2}) = \frac{1}{2}$ is definable (by $p \wedge \neg p$) but not distributive: $f(1 \sqcap 0) = \frac{1}{2} \neq 0 = f(1) \sqcap f(0)$

Moreover, the latter counterexample points at an insurmountable problem:

Theorem 1.13
The class of distributive functions is not definable by a set of connectives.

So, neither 3-valued functions which are distributive with respect to $\sqcap$, nor 4-valued functions with respect to $\sqcap$ or $\sqcup$, are in general definable by a set of connectives. The simple reason is that if they were, the standard connectives, being distributive, would be definable, and so would their compositions be. In other words, distributivity does not correspond to a closed class. By proposition 1.7 this implies the above theorem.

Hence distributivity is a powerful tool for producing extended truth tables, but, though appealing, it is not a proper condition on truth functions in general. Next we shall come across other conditions suggested in the literature.

other conditions proposed

Throughout the chapter we have noticed the use of conditions by other authors, for example, PERS (Fine, Blamey), CCLOS (van Benthem, Langholm), DUAL (Visser). But in fact some other conditions have been proposed as well, and we will treat them here.

In a modal setting, [Hu81] suggests a condition of *refinability* (besides PERS) to the effect that indeterminacies can be resolved both ways:

REF* If $[\![p]\!](s) = \frac{1}{2}$ then there exist s', s'' such that $s \sqsubseteq s'$, $s \sqsubseteq s''$, $[\![p]\!](s') = 1$ and $[\![p]\!](s'') = 0$.

This condition may be motivated by what Langholm calls *determinability*[41]: a gap means insufficient data. Transferred to our extensional semantics REF* is the condition *resolution* discussed in [Ve87].[42]

RES f is *resolute* if $f(\vec{x}) = \frac{1}{2}$ implies there exist $\vec{y}, \vec{z} \sqsupseteq \vec{x}$ such that $f(\vec{y}) = 1$ and $f(\vec{z}) = 0$.

Notice that RES differs only from Jaspars' reformulation of CCLOS mentioned in section 1.2 in that RES has 'and' where REF has 'or', a small difference with significant

[41] At least this is a way to interpret a remark in the introduction of [La88, p.3]; later on [La88, p.18] the technical appearance of determinability is that of CCLOS.

[42] In his account of 'supertruth' (cf. van Fraassen's supervaluations) [Fi75a, p.278/9] also discusses a condition called 'resolution' which however is completely equivalent to Humberstone's REF*, modulo persistence.

consequences. Veltman himself notices that resolution is only proper for certain f, interpreting standard connectives. Resolution can be related directly to distributivity: PERS + RES are equivalent to DISTR, modulo CCLOS. This can be used to show that RES is not definable in the sense of this chapter: for if it were, then, by proposition 1.5 and theorem 1.4, PERS + CCLOS + RES would be definable, and by the above equivalence, so would CCLOS + DISTR be. However, the counterexample to definability of DISTR (cf. theorem 1.13) is also CCLOS, which shows that CCLOS + DISTR is not a closed class, contradicting definability. Therefore the class of resolute functions is not definable.

Apart from existing conditions we may also think of new ones. Notice in this respect that PERS, DISTR and the like refer to the ordering of extension ($\sqsubseteq$). A striking possibility is to consider the $\leq$ ordering instead. For example, it can be shown that $\leq$–monotonicity forms a closed class, and seems to be defined by the *positive* formulas of 3– and 4–valued logic.

other connectives proposed

Also we might consider other connectives (or other interpretations) than the ones we have been focussing on. Again, without appeal to 'intrinsic naturalness', we can but inspect existing proposals. In fact we have to restrict ourselves to only some proposals, since on this side possibilities abound.

The first alternative concerns [Łu20]'s system Ł$_3$ which extends the standard system $\{\neg, \wedge\}$ with the stronger implication $\mapsto$ (see the following truth table). This is a proper extension of the standard system, since $\mapsto$ is neither persistent nor free and hence not definable by $\neg, \wedge$. However, $\mapsto$ *is* classically closed, and so in expressive force Ł$_3$ is a subsystem of $\{\neg, \wedge, \sim\}$.[43] The converse holds too, because of $\sim p = p \mapsto \neg p$. So Ł$_3$ just appears to be another way to define the class CCLOS! A similar result holds for the even stronger implication $\supset$, which is interdefinable with respect to $\sim$, modulo $\neg, \wedge$: $p \supset q = \neg(\neg \sim p \wedge \neg q)$ and $\sim p = p \supset \neg(p \supset p)$.

$\mapsto$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	$\frac{1}{2}$
0	1	1	1

$\supset$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	1	1	1
0	1	1	1

$\xrightarrow{w}$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
0	1	$\frac{1}{2}$	0

At least one other proposal deserves mentioning: the Bochvar/weak Kleene interpretations of the usual connectives, represented in the table by $\xrightarrow{w}$, which differ from their usual ('strong Kleene') counterparts by getting the value $\frac{1}{2}$ in each row and column headed by $\frac{1}{2}$. Of course the 'weak' tables still satisfy PERS, CCLOS and FREE, so it follows by theorem 1.5 that these new connectives can all be defined in terms of $\neg, \wedge$ (in their strong interpretations, of course).[44] Here the converse does not hold: $\wedge$ is not definable by, for example, $\neg$ and $\xrightarrow{w}$. This is a straightforward consequence of

[43] Łukasiewicz noticed that $\wedge$ is redundant: $p \wedge q = \neg((q \mapsto p) \mapsto \neg q)$.

[44] For example, weak conjunction as defined in the proof of theorem 1.5.

the fact that the weak connectives semantically map $3^n - 2^n$ onto $\frac{1}{2}$. This condition presumably describes the weak Kleene system.

Chapter 2

Modal definability

In this chapter we focus on the expressiveness of the modal language with respect to partial world semantics. Extending functional completeness from partial propositional logic to modal logic faces a number of problems. *A priori* it may not even be clear how to *define* functional completeness for (partial) modal logic. Such a definition is feasible and will be presented in section 2.2, which is dedicated to what may be called the *structural* approach to modal completeness. One essential observation is that the modal operator $\Diamond$ corresponds to the 'inverse image' operation R^{-1}, where R is the accessibility relation.[1] This is the most obvious counterpart of the kind of propositional definability studied in the first chapter. Here the leading question is: which formulas correspond to which semantic operations under what conditions? Though interesting, this method turns out to suffer from severe limitations.

Hence we will employ a different, *reductive* approach in section 2.4. There we do not seek for an (in some sense) absolute characterization, which relates syntax to semantics, but for a more moderate and relative characterization, which relates modal logic to first order predicate logic. For example, the first order formula $\exists y(Rxy \wedge Py \wedge Qy)$ corresponds to the modal formula $\Diamond(p \wedge q)$, since they are verified by essentially the same models. However, not every first order formula over R, P, Q corresponds to a modal formula. Then the definability question becomes: which modal formulas correspond to which first order formulas under what conditions?

Crucial preservation properties are introduced and studied in the intermediate section 2.3, including some other related techniques. We start with an introductory section which deals with the partial models that will be used to interpret the modal language.

2.1 Partial models for modal logic

At first sight we may wonder why modal definability is so much harder than propositional definability: after all, by the presence of the modal operators $\Box$ and $\Diamond$ we

[1] $R^{-1}[X] = \{y \mid \exists z \in X : Ryz\}$.

have extra expressive power. Yet this only holds for a truth-functional interpretation of modalities.[2] This truth-functional semantics has been treated in the previous chapter, so we cannot expect anything new from this perspective. Moreover, the truth-functional approach lacks the flexibility which is typical for modal logic, and in fact does not capture the basic intuition of necessity.

Therefore we adopt the much more flexible Kripke semantics[34] which is characterized by the presence of frames underlying the models. Recall that a *frame* is a pair $\langle S, R\rangle$, where S is a set of (as such unstructured) *situations* and $R \subseteq S \times S$ an *accessibility relation* between situations. A *Kripke model* M based on frame $\langle S, R\rangle$ is a triple $\langle S, R, V\rangle$, where V is an interpretation of atomic propositions, depending on the situation. In general V may be undefined or overdefined on the situations. If V is *classical*, i.e. always assigns either *true* or *false*, the situations are called *possible worlds*; if V is partially defined they are called *coherent situations* or *partial worlds* and if V is always defined but maybe overdefined, *total situations*.

The modal operators are interpreted relative to the accessibility relation; in possible worlds models we have: ($R[s] = \{t \mid sRt\}$, for convenience)

$$M, s \models \Box\varphi \Leftrightarrow \forall t \in R[s] : M, t \models \varphi$$
$$M, s \models \Diamond\varphi \Leftrightarrow \exists t \in R[s] : M, t \models \varphi$$

By contrast, in partial models we have separate truth and falsity relations, which we symbolize by $\models$ and $=\!\!|$, respectively.

The standard truth and falsity conditions for partial Kripke semantics are:[5]

$$M, s \models \Box\varphi \Leftrightarrow \forall t \in R[s] : M, t \models \varphi \quad M, s =\!\!| \Box\varphi \Leftrightarrow \exists t \in R[s] : M, t =\!\!| \varphi$$
$$M, s \models \Diamond\varphi \Leftrightarrow \exists t \in R[s] : M, t \models \varphi \quad M, s =\!\!| \Diamond\varphi \Leftrightarrow \forall t \in R[s] : M, t =\!\!| \varphi$$

We omit M if possible; $s \models \varphi$ should be read as 's supports (or, verifies) φ' or 'φ is true in s', and $s =\!\!| \varphi$ as 's rejects φ' or 'φ is false in s.

2.2 Modal truth functions

Perhaps the most evident approach to modal definability would be to extend propositional definability by incorporating the *global* effect of modal models into the notions. Kripke semantics is not truth-functional (in the usual sense of the word), since truth and falsity in one world may depend on truth and falsity in other worlds. As such this does not preclude the possibility of functional completeness; it just means that the semantic objects described by (syntactic) formulas will be more involved. What

[2] Łukasiewicz once advocated such an approach to necessity.

[3] This type of possible world semantics is usually attributed to Saul Kripke, who was one of the originators, although, for example, Stig Kanger and Jaakko Hintikka did pioneering research in this area.

[4] Though Kripke semantics is already quite flexible, we will study more general forms of possible world semantics in later chapters (viz. the chapters 4, 6, 7).

[5] From a 'partial' point of view, these are the most simple and obvious clauses one can think of, so it seems. We will, however, discuss alternative proposals in later chapters.

we need is what we will call *modal truth functions*, which take the set of worlds into account. This new perspective still leaves a lot of freedom for a precise definition of the notion of modal truth function. One of the essential parameters controlling such a notion concerns the amount and nature of the structural information we are willing to presuppose. Some options for this starting point are:

- a *designated* frame, i.e. a given frame with a fixed world in it;
- a given frame *simpliciter*;
- a class of frames (possibly fulfilling structural conditions);
- a class of models (idem).

These possibilities can obviously be mixed; in addition to this liberty, there are several ways to connect the presupposed structure to definability. For example, assuming a frame, we may require the interpreted formula to be equal to the modal truth function in *some* world or, by contrast, in *every* world. We opt for a given frame and global definability ('all worlds'), because this notion of *modal truth function* seems the most manageable and perspicacious one.

To introduce this notion observe that a formula can be interpreted in a Kripke frame as a set of possible worlds (its denotation). Consequently, a (modal) operator can be interpreted as a semantic operation on sets of possible worlds. For example, $[\![\Box]\!](X) = Y$ for some sets of worlds X, Y may correspond to the truth conditions of $\Box\varphi$, if $[\![\varphi]\!] = X$ and $[\![\Box\varphi]\!] = Y$.[6]

To generalize this to partial semantics, first notice the familiar equivalence of subsets and their characteristic functions, i.e. of $\wp(S)$ and 2^S. Switching to this functional format, and generalizing to k-valued interpretations ($k = 2, 3, 4$), a formula of partial modal logic may be conceived as a *frame valuation*, i.e. a mapping from situations to (k) truth values. Given some frame the interpretation of an arbitrary formula depends on the frame valuations assigned to the primitive propositions it contains. So the modal truth functions are operators on frame valuations. Formally, a modal truth function f is a function $f : (k^S)^n \longrightarrow k^S$. Bearing in mind that frame valuations are the modal counterparts of propositional truth values, we arrive at definitions of the basic notions *modal function space* and *modal definability*.[7]

Definition 2.1 (modal function space)
$\mathsf{F}_{S,k}$, *the* **modal function space** *for the k-valued case with set of situations S, stands for* $\bigcup_{n\in\omega}(k^S)^{(k^S)^n}$.

Instead of the valuation V, we use the equivalent notion of an (n-tuple) frame valuation $\vec{\xi}$, which is closer to the notation for propositional definability. V and $\vec{\xi}$ are related by:

[6] This construal is one way of putting the idea of neighbourhood semantics, as explained in chapter 6.

[7] The technical notions *projection*, *composition* and *closed class* (cf. the appendix to chapter 1) can be adapted likewise.

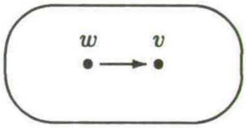
$$V(p_i, s) = \xi_i(s)$$

Let $[\![\cdot]\!]_{\vec{\xi}}$ recursively extend $[\![p_i]\!]_{\vec{\xi}} = \xi_i$, where $\xi_i \in k^S$.

Definition 2.2 (modal definability)
Let $\langle S, R \rangle$ be a fixed frame. A n-place modal truth function $f \in \mathsf{F}_{S,k}$ is **definable** *by a formula φ using the atoms $p_1, \ldots, p_n$ iff $[\![\varphi]\!]_{\vec{\xi}} = f(\vec{\xi})$. Likewise, a class $\mathsf{C} \subseteq \mathsf{F}_{S,k}$ is definable by a set of logical constants C iff*
(i) every $f \in \mathsf{C}^{(n)}$ is definable by some $\varphi \in \mathcal{L}_C\{p_1, \ldots, p_n\}$ and conversely
(ii) for all $\varphi \in \mathcal{L}_C\{p_1, \ldots, p_n\}$ the function $f : \vec{\xi} \mapsto [\![\varphi]\!]_{\vec{\xi}}$ is in $\mathsf{C}^{(n)}$.

The definition is sufficiently general to allow a renewed look on definability, even with respect to ordinary two-valued semantics. Then it shows that the expressive power of the ordinary modal language is very limited: for non-trivial frames there will be modal truth functions that do not correspond to any formula.

Example 2.1
Let $n = 1$, $k = 2$ and consider the displayed frame.



Not all operators are expressible since, roughly speaking, the fact that v does not 'see' w prohibits definability of operators in which the state-of-affairs in w influences the one in v. More technically, it can be shown by induction that for each formula $\varphi \in \mathcal{L}_{\neg,\wedge,\Box}\{p\}$, $[\![\varphi]\!]_\xi(v)$ only depends on $\xi(v)$, i.e. if $\xi(v) = \eta(v)$ then $[\![\varphi]\!]_\xi(v) = [\![\varphi]\!]_\eta(v)$. This leaves a lot of modal truth functions undefined, e.g. the function f provided by the following table:

ξ		$f(\xi)$	
w	v	w	v
1	1	0	1
1	0	0	0
0	1	0	0
0	0	0	0

A fortiori, the *extended modal* language $\mathcal{L}_{\neg,\wedge,\Box,\star,\sharp,\sim}(Prop)$ is not functionally complete with respect to 4-valued semantics either. Since we feel that the 'poor but proud austerity' of the standard modal operators should not be sacrificed too easily to more (e.g. bimodal) or stronger (e.g. dyadic) modalities, there is a need for suitable restrictions that may help to define the (extended) standard language.

One condition that comes to mind is *automorphism invariance*, which holds vacuously in the purely propositional case:

AUT Relative to a frame $\langle S, R\rangle$ a k-valued, n-ary modal truth function f is *automorphism invariant* iff $fa(\vec{\xi}) = af(\vec{\xi})$, for all automorphism a with respect to the given frame and all $\vec{\xi} \in (k^S)^n$.

Here the notion 'automorphism with respect to $\langle S, R\rangle$' is defined as usual: a is a *bijective* (1-1) function on S that respects R, i.e. $xRy \Rightarrow a(x)R\,a(y)$. Its use in the definition of 'AUT' in fact needs the hereditary effect on frame valuations: $a(\vec{\xi}) = \langle a(\xi_1), \ldots, a(\xi_n)\rangle$ where $a(\xi_i)$ is defined pointwise by $a(\xi_i) = \lambda s \xi_i(a(s))$.

It is easily verified that AUT holds for the extended modal language and in general it indeed restricts the semantic class, yet it does not succeed in characterizing the language. In some cases the restriction has no effect: reinspection of example 2.1 shows that the only automorphism on the frame is the identity mapping.

Without strong modal additions it seems hard to fully characterize the extended language. To get an idea of what is at stake, we focus on the somewhat simpler two-valued case, where a frame valuation is essentially a set of worlds. Observing what *is* expressible for the frame in example 2.1, we find evidence for a positive definability result.

Example 2.1 (continued)
Note that all expressible unary functions are unions of the truth functions related to the formulas in the table:

p		$\Diamond\Diamond p$		$p \wedge \Diamond p$		$p \wedge \Diamond\neg p$		$\neg p \wedge \Diamond p$		$\neg p \wedge \Diamond\neg p$		$p \wedge \Box\Diamond\neg p$		$\neg p \wedge \Box\Diamond\neg p$	
w	v	w	v	w	v	w	v	w	v	w	v	w	v	w	v
1	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1

Notice that the expressible operators 'run in blinkers': the operation transferring X into $R^{-1}[X]$ is expressible (by $\Diamond$), its inverse $X \mapsto R[X]$ may not be expressible.[8] To illustrate this we return to the example once more. (Let ξ be the characteristic function of a set $X \subseteq \{w, v\}$.)

Example 2.1 (second continuation)
It may help to contrast $R[X]$ and $R^{-1}[X]$:

X		$R[X]$		$R^{-1}[X]$	
w	v	w	v	w	v
1	1	0	1	1	0
1	0	0	1	0	0
0	1	0	0	1	0
0	0	0	0	0	0

So, $R^{-1}[\cdot]$ is expressed $\Diamond p$, $R[\cdot]$ is not expressible, as follows from earlier observations.

[8] Recall that $R^{-1}[X] = \{x \mid \exists y \in X : xRy\}$ and $R[X] = \{y \mid \exists x \in X : xRy\}$.

In temporal logic the problem of one-sided definability is solved by having operators looking both ways (e.g. **F** and **P**). A good illustration of structural definability within tense logic can be found in [vB86, p.102/3]. Van Benthem introduces a notion of *continuity* (i.e. $\bigvee$–distributivity):

CONT ($k = 2, n = 1$) On a frame $\langle S, R\rangle$ a unary modal truth function f is *continuous* iff $f(\bigcup_i X_i) = \bigcup_i f(X_i)$, for all families $\{X_i \mid i \in I\}$ where $X_i \subseteq S$.

Now let $\langle \Re, <\rangle$ be the set of reals ordered by 'smaller than'.

Theorem 2.1 *For the frame $\langle \Re, <\rangle$, AUT $\cap$ CONT$^{(1)}$, the set of unary modal truth functions which are automorphism invariant and continuous, is definable by $\mathcal{L}_{\perp,\vee,\mathbf{F},\mathbf{P}}\{p\}$.*

Setting aside such bimodal definability, it will be hard to find equally informative results. The point is that there is no *a priori* difference between relations and their inverses, whence it is impossible to give a semantic property possessed by R, but not by R^{-1}. The only way out may be a direct invocation of R^{-1}. Then the result for the classical case would amount to

Theorem 2.2 *For Kripke semantics, the language $\mathcal{L}_{\neg,\wedge,\Box}(Prop)$ defines the class Λ_2 which contains the projections and is closed under the operations complementation, intersection, $R^{-1}[\cdot]$ and composition.*

Of course this is nothing but an algebraic reformulation of the syntactic structures of modal formulas within semantics; nevertheless it is the best we can do in general, under the present perspective.

Similar results are readily obtained in a partial setting, once R^{-1} has been recast as the operation ϱ, which interprets $\Diamond$.

Definition 2.3
For a frame $\langle S, R\rangle$ the unary operation $\varrho : k^S \longrightarrow k^S$ is defined by ($k = 2, 3$)

$$\varrho(\xi)(s) = \begin{cases} 1 & \text{iff } sRt \ \& \ \xi(t) = 1 \text{ for some } t \\ 0 & \text{iff } \xi(t) = 0 \text{ for all } t \text{ such that } sRt \\ \frac{1}{2} & \text{else} \end{cases}$$

Thus we have $(\varrho\xi)^{-1}[1] = R^{-1}[\xi^{-1}[1]]$ and $(\varrho\xi)^{-1}[0] = (R^{-1}[\xi^{-1}[0]^c])^c$. Observe that the definition can be simplified to

$$\varrho(\xi)(s) = \max_{t\in R[s]} \xi(t),$$

which on its turn can be generalized (for $k = 2, 3, 4$) to

$$\varrho(\xi)(s) = \bigvee_{t\in R[s]} \xi(t).$$

This reformulation shows that ϱ preserves disjunctions. Thus we obtain a similar algebraic characterization, featuring operators which are counterparts to the connectives; e.g. $-$ for $\neg$, $\times$ for $\wedge$, $\frac{1}{2}$ for $\star$, 2 for $\sharp$, ϱ for $\Diamond$ and, say, $\ominus$ for $\sim$, etcetera.

Theorem 2.3 **(Λ_k)** *For the trivalent (quadrivalent) semantics, the respective language $\mathcal{L}_{\neg,\wedge,\Box,\star,\sim}(Prop)$ ($\mathcal{L}_{\neg,\wedge,\Box,\star,\sim,\sharp}(Prop)$) defines the class Λ_3 (Λ_4) which contains the projections, $\frac{1}{2}$ (and 2) and is closed under the operations $-$, $\times$, ϱ, $\ominus$ and composition.*

Having given this rather crude form of definability we may impose the conditions proposed in chapter 1, appropriately generalized to the modal case (by pointwise definition). This amounts to replacing (vectored) truth values by (vectored) frame valuations in the definitions of PERS, CCLOS etcetera. Then these conditions combined with Λ_k are definable by the subsets of the extended modal language that one would expect from the propositional case. For example, PERS $\cap\Lambda_3$ is definable by $\neg, \wedge, \star, \top, \Box$.

Although this line may be worth pursuing, we will not do so here. Instead we will pay more attention to the full extended language and continue with a less revolutionary approach, using a translation of modal logic into the first order predicate calculus. This also involves the introduction of an important equivalence relation between models, which will be defined in the next section.

2.3 Preservation under bisimulations

One way to characterize normal modal logic is in terms of its truth preservation under semantic operations such as filtration[9], disjoint union, generated submodels and p-morphisms. In classical modal logic these notions are covered by the more general notion of *bisimulation*. In this section we will try and see whether a similar characterization can be obtained for partial modal logic.

To have an equivalence relation between models that, in some sense, characterizes them as being models of modal logic, the notion of *bisimulation* has been suggested.[10] We accommodate this notion for partial semantics:

Definition 2.4 (bisimulation)
A relation $Z \subseteq S \times S'$ is a **bisimulation** *between two models $\langle S, R, V\rangle$ and $\langle S', R', V'\rangle$ iff*[11]

1. $(sZs' \ \&\ sRt) \Rightarrow \exists t' : (tZt' \ \&\ s'R't')$
2. $(sZs' \ \&\ s'R't') \Rightarrow \exists t : (tZt' \ \&\ sRt)$

[9]Filtrations will be considered in chapter 4, where we derive the finite model property for several systems of partial modal logic.

[10][vB85] uses the term 'p-relation', [vB84b] 'zigzag relation' (both are 'full'), [vB90] 'bisimulation' (without bitotality). 'Bisimulation' is independently used in computer science, see e.g. [St87].

[11]Without (3) Z is a bisimulation between frames. Notice that (1) and (2) can be summarized by: ($M = M_1, M' = M_2$, etcetera, $s_i, t_i \in S_i$, $i \neq j$, $i, j = 1, 2$) $(s_1 Z s_2 \ \&\ s_i R_i t_i) \Rightarrow \exists t_j : (t_1 Z t_2 \ \&\ s_j R_j t_j)$, cf. [dR90].

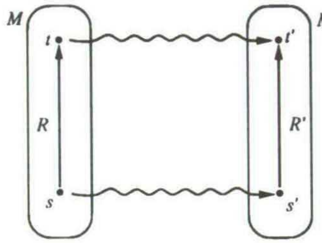
3. $V(p, s) = V'(p, s')$ *if* sZs'

Z is a **full** *bisimulation when in addition to (1-3) clause (4) holds:*

4. $\mathrm{dom}(Z) = S$ *and* $\mathrm{ran}(Z) = S'$.

This definition may be clarified by the following diagram.

Figure 2.1: *Bisimulation diagram*



A bisimulation (or the existence of a suitable bisimulation) is indeed an equivalence relation, both between global models and between frames (or between local models and frames), symbolized by $M \bowtie M'$ and $F \bowtie F'$ ($M, s \bowtie M', s'$ and $F, s \bowtie F', s'$ respectively): the identity mapping guarantees reflexivity, the inverse relation symmetry, and composition transitivity.

The obvious question now emerging is whether partial truth and falsity are preserved under bisimulations, like classical truth is. The answer is 'yes', even when extra operators $\top, \bot, \star, \natural, \sim, \partial$ are added to the modal language.

From now on, let the *extended* modal language be the propositional language which contains the above operators and $\Box$ and $\Diamond$, i.e. essentially[12], $\mathcal{L}_{\neg,\wedge,\Box,\sim,\star,\natural}(Prop)$. We also make use of a partialized notion of *modal equivalence*: (likewise for *propositional, elementary* equivalence)

$$M, s \overset{\text{mod}}{\equiv} M', s' \quad \text{iff} \quad M, s \models \varphi \Leftrightarrow M', s' \models \varphi \text{ for all modal } \varphi.$$

Then

$$M, s \bowtie M', s' \Rightarrow M, s \overset{\text{mod}}{\equiv} M', s'.$$

as is shown by the following proposition.

Proposition 2.1 (preservation under bisimulations)
If Z is a bisimulation between M and M', $M, s \models \varphi \Leftrightarrow M', s' \models \varphi$, for all φ and s, s' such that sZs'.

[12]Cf. chapter 1, theorems 1.2 and 1.6; of course $\natural$ only pops up in the 4-valued case.

Proof: by simultaneous induction on the structure of φ we prove that

$$M, s \models \varphi \Leftrightarrow M', s' \models \varphi \quad \& \quad M, s \mathrel{=\!\!|} \varphi \Leftrightarrow M', s' \mathrel{=\!\!|} \varphi$$

for all s, s' such that sZs'. For the propositional variables this is given by definition, for the 0-place constants it is obvious, and for the other connectives it is easy. We will therefore only spell out the case of $\sim$, as well as the crucial modal step. So assume that the properties above hold for φ (IH), let sZs' then:

- $M, s \models \sim\varphi \Leftrightarrow M, s \not\models \varphi \Leftrightarrow$ (IH) $M', s' \not\models \varphi \Leftrightarrow M', s' \models \sim\varphi$.
- $M, s \mathrel{=\!\!|} \sim\varphi \Leftrightarrow M, s \models \varphi \Leftrightarrow$ (IH) $M', s' \models \varphi \Leftrightarrow M', s' \mathrel{=\!\!|} \sim\varphi$.
- $M, s \models \Box\varphi \Leftrightarrow \forall t \in R[s] : M, t \models \varphi \overset{*}{\Leftrightarrow} \forall t' \in R'[s'] : M', t' \models \varphi \Leftrightarrow M', s' \models \Box\varphi$, where

  $\overset{*}{\Rightarrow}$: assume the left-hand side, then for an arbitrary $t' \in S'$ by clause (2) of bisimulation there is a $t'' \in S$ such that $t''Zt'$ and sRt'', so $M, t'' \models \varphi$, and so by IH: $M', t' \models \varphi$;

  $\overset{*}{\Leftarrow}$: by a symmetric argument, using clause (1) of the definition of bisimulation.

- $M, s \mathrel{=\!\!|} \Box\varphi \Leftrightarrow \exists t \in R[s] : M, t \mathrel{=\!\!|} \varphi \overset{*}{\Leftrightarrow} \exists t' \in R'[t'] : M', t' \mathrel{=\!\!|} \varphi \Leftrightarrow M', s' \mathrel{=\!\!|} \Box\varphi$, where

  $\overset{*}{\Rightarrow}$: assuming the left-hand side, by (1) there is a $t' \in S'$ such that tZt' and $s'Rt'$, so by IH: $M', t' \mathrel{=\!\!|} \varphi$;

  $\overset{*}{\Leftarrow}$: analogously, by (2).

■

Generally speaking, the notion of bisimulation has turned out to be strikingly stable and applicable to many branches of modal logic. This is reflected in partial semantics by the fact that additional connectives (if extensional) do not mar the game. In other words, differently, the preservation holds for the extended, functionally complete propositional language with modal operators added.

An older characteristic of normal modal logic is truth preservation under p-morphisms.[13] From our modern perspective a p-morphism is essentially a bisimulation which is a *function*, often assumed to be surjective (onto).

Another simple but useful result in classical modal logic is the so-called *generation lemma*. Informally speaking, it says that the truth value of a formula in some world only depends on that particular world and those worlds accessible from it in a finite number of steps. This result is easily transferred to partial modal logic.

Definition 2.5 (generated submodel)
Let $M = \langle S, R, V\rangle$ be a partial modal model. A model $M' = \langle S', R', V'\rangle$ is a generated submodel of M iff

- $S' \subseteq S$;

[13] From [Se70], but it has presumably been folklore for quite a while; the term is said to be an abbreviation of 'pseudo-epimorphism'.

- *if $s \in S'$ and sRt then $t \in S'$;* (*i.e.* $R[S'] \subseteq S'$)
- *$sR't$ iff $s, t \in S'$ and sRt;* (*i.e.* $R' = R \cap S' \times S'$)
- $V'(p, s) = V(p, s)$ *if* $s \in S'$.

Notice the relation *generated submodel* is a special case of *bisimulation*: for if $S' \subseteq S$, let Z be the embedding of S' in S (i.e. the identity function restricted to S'). So again we obtain an easy consequence of proposition 2.1.

Corollary 2.1 (generation lemma)
If M' is a generated submodel of M, then for all s in M': $M, s \models \varphi \Leftrightarrow M', s \models \varphi$.

An important class of generated submodels are the so-called *rooted* submodels, in which there is a world (situation) (the *root* or *generator*) from which all worlds of the generated submodel are accessible in a finite number of jumps, i.e. the submodel is generated from the root. First we define the iterative composition of a relation.

Definition 2.6 (iterative relational composition)
If R is an arbitrary relation on some set S, let

- $R^0 = \{\langle s, s\rangle \mid s \in S\}$;
- $R^{n+1} = R \bullet R^n = \{\langle s, t\rangle \mid \exists s' \in S : \langle s, s'\rangle \in R \,\&\, \langle s', t\rangle \in R^n\}$;
- $R^* = \bigcup_{n\in\omega} R^n$.

Definition 2.7 (rooted submodel)
$M_s = \langle S_s, R', V'\rangle$ is a rooted submodel of $M = \langle S, R, V\rangle$ (or a submodel generated by s) iff M_s is a generated submodel of M where $S_s = R^*[s]$.

The definition is justified by the fact that S_s is closed with respect to R. So we have a constructive kind of generated submodel at our disposal.

Corollary 2.2 *The generation lemma holds for M_s.*

This more particular form of the generation lemma can be used to derive another familiar property of modal models.[14] M is a local model for φ if for some s in M: $M, s \models \varphi$ (we sometimes also say that $\langle M, s\rangle$ is a local model); M is a global model for φ if for all s in M: $M, s \models \varphi$.

Corollary 2.3 (preservation under disjoint union)
If φ is locally (globally) true for some family of models, it is also locally (globally) true for the disjoint union of those models.

[14] The special form of the generation lemma and preservation under disjoint union are intensively used in part III of this thesis. Since these cases are subsumed under bisimulation there is virtually no technical reason to discuss them separately, but we consider these special instances to be more transparant and easier to visualize.

Because of preservation and flexibility one may cherish the hope that bisimulation techniques may be applicable to definability matters in partial modal logic too, just like in 'classical' logic. Before realizing this expectation we are eager to make some simple, yet useful observations. The first is that proposition 2.1 can be converted in the important special case where the models are finite.

Proposition 2.2 *For finite models modal equivalence entails bisimulation equivalence (in the extended language):*

$$M, s \overset{\text{mod}}{\equiv} M', s' \Rightarrow M, s \bowtie M', s'.$$

Proof: for given models M and M' define the relation Z by $tZt' \Leftrightarrow M, t \overset{\text{mod}}{\equiv} M', t'$. Then Z is a bisimulation between M and M' since

1. Let (i) tZt' and (ii) tRu. We have to show that uZu' and $t'R'u'$ for some u'. Suppose on the contrary that $R'[t'] \cap Z[u] = \emptyset$. Thus for all $v \in R'[t']$ there is an extended modal formula φ_v such that either $M, u \models \varphi_v$ & $M', v \not\models \varphi_v$ or $M, u \not\models \varphi_v$ & $M', v \models \varphi_v$. Now define ψ_v by $\psi_v = \varphi_v$ for the first possibility, and $\psi_v = \sim\varphi_v$ for the second one. Then in either case $M, u \models \psi_v$ & $M', v \not\models \psi_v$. Since $R'[t']$ is finite the formula

$$\psi = \bigwedge_{v \in R'[t']} \psi_v$$

is well-defined and moreover (iii) $M, u \models \psi$ and (iv) $M', v \not\models \psi$ for all $v \in R'[t']$. By (ii) and (iii) $M, t \models \Diamond\psi$ and so by (i) and the definition of Z: $M', t' \models \Diamond\psi$, therefore $M', v \models \psi$ for some $v \in R'[t']$, contradicting (iv).

2. The reverse clause goes in the same fashion.

3. If tZt' then $M, t \models \varphi \Leftrightarrow M', t' \models \varphi$ for all (extended modal) φ. Now inspection of the cases for the literals ($\varphi = p$ or $\varphi = \neg p$) shows that $V(p, t) = V'(p, t')$.

So, if $M, s \overset{\text{mod}}{\equiv} M', s'$ then by definition of Z sZs', and so $M, s \bowtie M', s'$. ∎

The second remark is that due to the global nature of modal equivalence we can stick to just *truth* preservation under bisimulations, without keeping track of the *falsity* side, which makes things a lot easier when trying to establish definability.

One way to make this approach to definability work is to translate the modal logic into first order predicate logic, and see what makes these translations special among the other first order formulas. The semantic property that captures the translations should not be surprising anymore.

2.4 Translating into first order logic

The idea behind the translational approach is to regard possible world models as a special kind of first order models where the objects are the worlds. Then one way to characterize classical modal logic is to use a formalization of the meta-language, which, by the nature of the truth conditions, amounts to a translation into first order logic. Apart from quantifiers and identity, the language of this first order logic contains only one dyadic predicate (to represent the accessibility relation), and one monadic predicate for each propositional variable of the original modal object-language. The definability question then becomes: which first order formulas express modal truth conditions? Depending on the art of the translation it is possible to characterize this sublanguage recursively. More interesting, however, and independent of the specific translation, is a purely semantic characterization: in [vB85] it is shown that what makes the translations special is their invariance under full bisimulations and generated submodels. This criterion has been simplified in [vB90] to preservation under *arbitrary bisimulations*. The main theorem of this section is a similar definability result for partial modal logic:

> *An (extended) partial first order formula $\varphi(x)$ with only monadic predicates and one special dyadic relation is equivalent to an (extended) partial modal formula iff φ is truth preserved under bisimulation.*

Here the *standard* partial first order language contains the operators $\neg, \wedge, \forall$, monadic predicates $P_1, \ldots, P_n$ and exactly one bivalently interpreted relation symbol R; the corresponding *standard* partial modal language contains $\Box$ instead of $\forall$ and R, and p_i instead of P_i. The *extended* partial (first order or modal) languages contain the additional operators $\star, \sim$, and in the quadrivalent case $\sharp$. The *persistent* languages are obtained from the extended ones by removing $\sim$ from the stock of logical constants. Finally the *positive* partial first order language contains the operators $\wedge, \vee, \forall, \exists, \star, \sharp$ (and $R, \vec{P}$, as always) whereas the *positive* partial modal language makes use of only the operators $\top, \bot, \wedge, \vee, \Box, \Diamond, \star, \sharp$ (and $\vec{p}$, of course). So, in an ultimate attempt to keep the terminology manageable, we dropped the attribute *propositional* from the modal languages.

The main results in this section will be demonstrated by a number of successive translations, which essentially reduce the problem of partial (relative) definability to that of classical (relative) definability.

From modal logic to first order logic

First consider a suitable translation from the language of classical modal logic into that of classical predicate logic. The standard translation ST ignores all connectives and merely changes the atoms and modal operators: (x is a given variable, P a distinct (monadic) predicate symbol for each propositional atom p, and R a fixed dyadic predicate symbol)

$$\begin{array}{ll} \mathrm{ST}(p) = Px & \mathrm{ST}(\neg\varphi) = \neg\mathrm{ST}(\varphi) \\ \mathrm{ST}(\varphi \wedge \psi) = (\mathrm{ST}(\varphi) \wedge \mathrm{ST}(\psi)) & \mathrm{ST}(\Box\varphi) = \forall y(Rxy \rightarrow [y/x]\mathrm{ST}(\varphi)) \end{array}$$

Then, a Kripke structure can be regarded as both a predicate logical model and a modal model:

$$M, w \models \varphi \quad \text{iff} \quad M \models \mathrm{ST}(\varphi)[w]$$

The classical result, from [vB85] and [vB90], shows what makes the formulas produced by the standard translation special among the arbitrary formulas in the first order language $\mathcal{L}_{\neg,\wedge,\forall}\{R, \vec{P}\}$: bisimulation preservation.

Theorem 2.4 (van Benthem)
A classical first order formula φ using $R, P_1, \ldots, P_n$ and exactly one free variable x is equivalent to a modal formula iff φ is preserved under bisimulation.

We are looking for similar 'partial' results in the rest of this section. We will not give direct proofs, since this would imply a thorough treatment of partial predicate logic, which goes beyond the scope of this thesis. Rather we wish to *reduce* the problem of determination by preservation under bisimulations between partial models to that between classical models. This involves the following translations:

- The standard translation ST from a classical(ly interpreted) modal language $\mathcal{L}_\Box$ to a classical first order language $\mathcal{L}_\forall$;[15]
- A similar translation, also denoted by ST, from a partial modal language $\mathcal{L}_{\sim,\Box}$ to a partial first order language $\mathcal{L}_{\sim,\forall}$;[16]
- A 'forward' translation $^+$ from a partial first order language $\mathcal{L}_{\sim,\forall}$ to a classical first order language $\mathcal{L}_\forall$;[17]
- A 'backward' translation $^\times$ from a classical modal language $\mathcal{L}_\Box$ to a partial modal language $\mathcal{L}_{\sim,\Box}$.

The most important translational steps are summarized in the next diagram:

Figure 2.2: *translational diagram*

$$\begin{tikzcd}
\mathcal{L}_\Box \arrow[r, "\mathrm{ST}"] \arrow[d, "\times"'] & \mathcal{L}_\forall \\
\mathcal{L}_{\sim,\Box} \arrow[r, "\mathrm{ST}"] & \mathcal{L}_{\sim,\forall} \arrow[u, "+"']
\end{tikzcd}$$

Then the main argument boils down to:

1. an extended first order formula φ is preserved under 'partial' bisimulations $\Leftrightarrow$

[15] To keep this survey transparant, we omit the similarity type (the non-logical vocabulary) from the specifications of these languages, as well as $\neg$ and $\wedge$.

[16] For partial logics, extend ST with obvious clauses, such as $\mathrm{ST}(\sim\varphi) = \sim\mathrm{ST}(\varphi)$, etcetera.

[17] This translation invokes an auxiliary translation $^-$.

2. its translation φ^+ is preserved under classical bisimulations $\Leftrightarrow$

3. there is an ordinary modal formula ψ which is classically equivalent to (validated by the same classical models as) φ^+ $\Leftrightarrow$

4. φ is equivalent to the extended modal formula $\psi^\times$.

From partial logic to classical logic, and back

Following the initial idea of [Gi74] for partial set theory, [La88] translates partial (propositional and predicate) logic into classical (propositional and predicate) logic. In this way he derives a number of definability results.[18] Let us first deal with the propositional languages. The trick is that for each atom p of the original language there are two counterparts p^+, p^- in the new language, where the (total) truth of p^+ and p^- amounts to the (partial) truth and falsity of p, respectively. Simplifying the presentation, the translation can be defined by recursively extending $^+$ and $^-$:

$$\begin{array}{ll}
\star^+ = \bot & \star^- = \bot \\
\natural^+ = \top & \natural^- = \top \\
(\neg\varphi)^+ = \varphi^- & (\neg\varphi)^- = \varphi^+ \\
(\varphi \wedge \psi)^+ = (\varphi^+ \wedge \psi^+) & (\varphi \wedge \psi)^- = (\varphi^- \vee \psi^-) \\
(\sim\varphi)^+ = \neg\varphi^+ & (\sim\varphi)^- = \varphi^+
\end{array}$$

By virtue of the other definitions we obtain, for example, $\top^+ = \top$, $\top^- = \bot$, $(p \vee \neg p)^+ = p^+ \vee p^-$, $(\neg(p \wedge \neg p) \vee q)^+ = p^- \vee p^+ \vee q^+$. Notice that for formulas that do not contain $\sim$, the translation produces a positive normal form. Also notice that $\bot$ can be eliminated by taking $p^+ \wedge \neg p^+$ instead, and $\top$ on its turn by $\neg\bot$, and finally $\vee$ and $\Diamond$ can be defined in terms of $\neg, \wedge, \Box$ by familiar equations. However, these replacements destroy a proper back and forth translation, and therefore can only be applied after the translation procedure.

We can extend this approach to modal formulas by putting:

$$(\Box\varphi)^+ = \Box\varphi^+ \quad (\Box\varphi)^- = \Diamond\varphi^-$$

This choice allows for a relative characterization of extended modal formulas. The following facts are expansions of results reported by Langholm:[19]

Lemma 2.1 (forward translation)
$^+$ *transforms extended modal formulas in* **standard** *modal formulas; moreover,* $\sim$*-free formulas are translated in* **positive** *formulas.*

Proof: In fact the same result holds for the 'negative' translation $^-$. That the output of both translations is of the intended format can be shown by a simultaneous induction on the structure of input-formulas. ■

[18] This alternative to our earlier method of is especially useful for partial *predicate* logic.

[19] [La88, p.22,23,26,27,111]

For a proof of our main theorem we also need a reverse function $^\times$:

$$\begin{array}{ll} (p^+)^\times = p & (p^-)^\times = \neg p \\ \top^\times = \sharp & \bot^\times = \star \\ (\varphi \wedge \psi)^\times = (\varphi^\times \wedge \psi^\times) & (\varphi \vee \psi)^\times = \neg(\neg\varphi^\times \wedge \neg\psi^\times) \\ (\neg\varphi)^\times = \sim\varphi^\times & \\ (\Box\varphi)^\times = \Box\varphi^\times & (\Diamond\varphi)^\times = \neg\Box\neg\varphi^\times \end{array}$$

Lemma 2.2 (backward translation)
$^\times$ *transforms standard modal formulas in* **extended** *formulas; $\neg$-free formulas are translated in* **persistent** *formulas. Moreover,* $(\varphi^\times)^+ = \varphi$, $[\![(\varphi^+)^\times]\!] = [\![\varphi]\!]$ *and* $[\![(\varphi^-)^\times]\!] = [\![\neg\varphi]\!]$.

Proof: straightforward induction ■

So, unlike $(\varphi^\times)^+$ and φ, $(\varphi^+)^\times$ and φ may not be identical, yet they are equivalent. These lemmas obviate the following proposition.

Proposition 2.3 (encoding partiality)
The mapping $^+$ is onto (and so is $^-$), i.e. [20]

$$[\mathcal{L}_{\neg,\wedge,\Box,\sim,\star,\sharp}(Prop)]^+ = \mathcal{L}_{\top,\bot,\neg,\wedge,\vee,\Box,\Diamond}([Prop]^\pm)$$
$$[\mathcal{L}_{\neg,\wedge,\Box,\star,\sharp}(Prop)]^+ = \mathcal{L}_{\top,\bot,\wedge,\vee,\Box,\Diamond}([Prop]^\pm).$$

Proof: The first lemma ensures that the functions map into the correct language, the second one that the functions are onto, since $(\varphi^\times)^+ = \varphi$, thus also $(\neg(\varphi)^\times)^- = \varphi$. ■

The function $^\times$ is strictly speaking not onto, but it is up to logical equivalence. Let us therefore turn to the semantic side of the translations.

There is an 1-1 correspondence between (general) partial modal models on the one hand and classical Kripke models on the other hand.

Proposition 2.4 (forward modal truth transfer)
For any general model $M = \langle S, R, V\rangle$ with $V : Prop \times S \longrightarrow 4$ define the corresponding classical model $M^* = \langle S, R, V^*\rangle$ by $V^*(p^+, s) = 1 \Leftrightarrow 1 \sqsubseteq V(p, s)$ and $V^*(p^-, s) = 1 \Leftrightarrow 0 \sqsubseteq V(p, s)$. Then for all (extended) modal formulas φ: $M, s \models \varphi$ iff $M^*, s \models \varphi^+$.

Proof: We can show by simultaneous induction that $M, s \models \varphi \Leftrightarrow M^*, s \models \varphi^+$ and $M, s \,=\!\!\!|\, \varphi \Leftrightarrow M^*, s \models \varphi^-$. For example, assuming the IH for an arbitrary φ, here are the induction steps for $\sim$ and $\Box$:

- $M, s \models \sim\varphi \Leftrightarrow M, s \not\models \varphi \Leftrightarrow$ (IH) $M^*, s \not\models \varphi^+ \Leftrightarrow M^*, s \models \neg\varphi^+ \Leftrightarrow$ $M^*, s \models (\sim\varphi)^+$
- $M, s \,=\!\!\!|\, \sim\varphi \Leftrightarrow M, s \models \varphi \Leftrightarrow$ (IH) $M^*, s \models \varphi^+ \Leftrightarrow M^*, s \models (\sim\varphi)^-$

[20] $[\Phi]^+ = \{\varphi^+ \mid \varphi \in \Phi\}$ and similarly for $[\Phi]^-$. Finally, $[\Phi]^\pm = \Phi^+ \cup \Phi^-$.

- $M, s \models \Box\varphi \Leftrightarrow \forall t \in R[s] : M, t \models \varphi \Leftrightarrow$ (IH) $\forall t \in R[s] : M^*, t \models \varphi^+ \Leftrightarrow$ $M^*, s \models \Box\varphi^+ \Leftrightarrow M^*, s \models (\Box\varphi)^+$
- $M, s \dashv \Box\varphi \Leftrightarrow \exists t \in R[s] : M, t \dashv \varphi \Leftrightarrow$ (IH) $\exists t \in R[s] : M^*, t \models \varphi^- \Leftrightarrow$ $M^*, s \models \Diamond\varphi^- \Leftrightarrow M^*, s \models (\Box\varphi)^-$

■

This result has a valid counterpart in predicate logic, which will also be of use for the special purpose language over R and $\vec{P}$. First extend the $^+$ translation to quantifiers and predicates:

$$(P\vec{x})^+ = P^+\vec{x} \qquad (P\vec{x})^- = P^-\vec{x}$$
$$(\forall x\varphi)^+ = \forall x\varphi^+ \quad (\forall x\varphi)^- = \exists x\varphi^-$$

Proposition 2.5 (forward first-order truth transfer)
For a general first order model $M = \langle S, V\rangle$ with[21] $V(P, s) \in 4$ define the classical model $M^* = \langle S, V^*\rangle$ by a similar transformation (replacing p in the previous proposition by P) Then for all (extended) first order formulas φ:

$$M \models \varphi[\vec{s}] \quad \textit{iff} \quad M^* \models \varphi^+[\vec{s}]$$

Proof: by an analogous induction. ■

Proposition 2.6 (backward truth transfer)
Under the same circumstances as in the previous propositions, for all standard modal formula φ: (similarly for first order formulas)

$$M^*, s \models \varphi \quad \textit{iff} \quad M, s \models \varphi^\times$$

Proof: follows immediately from lemma 2.2 and proposition 2.4. ■

(There is an obvious inverse to the model transformation, leading from a classical model N to a partial, possibly incoherent model N', cf. the construction in proposition 2.4.)

partial and classical bisimulations

A final step for reducing the (relative) definability problem for partial modal logic to that for classical modal logic involves a direct relation between partial bisimulation (here in the sense of bisimulations between partial models) and total bisimulations. In esoteric terms, bisimulations simulate bisimulations.

Proposition 2.7 (partial vs total bisimulations)
An extended first order formula φ over $\vec{P}$ and (bivalently interpreted) R is truth preserved under partial bisimulations iff φ^+ is preserved under classical bisimulations.

[21] V is usually presented by pairs $\langle V(P)^+, V(P)^-\rangle$ marking its positive and negative denotation. The model transformation may then be defined as $V^*(P^+) = V(P)^+$ and $V^*(P^-) = V(P)^-$, which is concise by possibly confusing.

Proof: there are two things which have to be shown:
($\Rightarrow$) Assume that partial bisimulations preserve the truth of φ. Suppose that Z is a bisimulation between the classical models M_1 and M_2, where $M_i = \langle S_i, V_i\rangle$ interprets the standard first order language with predicate symbols $R, P_1^+, P_1^-, \dots, P_n^+, P_n^-$ and $V_i(R) = \mathsf{R}_i$.[22] Define the partial models $M_i' = \langle S_i, V_i'\rangle$ by $V_i'(R) = \langle \mathsf{R}_i, \mathsf{R}_i^c\rangle$ and $V_i'(P_j) = \langle V_i(P_j^+), V_i(P_j^-)\rangle$. Z also relates M_1' and M_2'; with respect to the latter models call it Z'. Then Z' is also a bisimulation: the structural conditions are met since $V_i(R)$ and $V_i'(R)$ amount to the same; furthermore $V_1'(P_j)^+ = V_1(P_j^+) \overset{Z}{=} V_2(P_j^+) = V_2'(P_j)^+$ and likewise $V_1'(P_j)^- = V_2'(P_j)^-$. So, since Z' is a bisimulation, Z' preserves φ and $M_i = M_i'^*$, we have that for all s_1, s_2 such that $s_1 Z s_2$: $M_1 \models \varphi^+[s_1] \Leftrightarrow$ (proposition 2.5) $M_1' \models \varphi[s_1] \Leftrightarrow M_2' \models \varphi[s_2] \Leftrightarrow$ (proposition 2.5) $M_2 \models \varphi^+[s_2]$ and so φ^+ is invariant under classical bisimulations.
($\Leftarrow$) By a very similar argument, now starting with the assumption of preservation of φ^+ under classical bisimulations and then considering a partial bisimulation Z between M_1 and M_2. Transforming them into Z^*, M_1^* and M_2^* (cf. proposition 2.5), we again arrive at truth preservation of φ under partial bisimulations. ∎

All the machinery needed for the definability theorem has now been arranged. Let $\varphi(x)$ be a first order formula from the extended language with one free variable x.

Theorem 2.5 (partial first order definability)
An extended partial first order formula $\varphi(x)$ with monadic predicates $P_1, \dots P_n$, and a dyadic, bivalently interpreted R is equivalent to an (extended) partial modal formula over $p_1, \dots, p_n$ iff φ is truth preserved under (partial) bisimulation.

Proof: $\varphi \in \mathcal{L}_{\star,!,\sim,\forall}(R, \vec{P})$ is truth preserved under bisimulations on partial models $\Leftrightarrow$ (proposition 2.7) its translation $\varphi^+ \in \mathcal{L}_\forall(R, \vec{P^+}, \vec{P^-})$ is preserved under bisimulation on classical models $\Leftrightarrow$ (theorem 2.4) φ^+ is equivalent to a modal formula $\psi \in \mathcal{L}_\Box(\vec{p^+}, \vec{p^-})$ $\Leftrightarrow$ (proposition 2.6) φ is equivalent to the modal formula $\psi^\times \in \mathcal{L}_{\star,!,\sim,\Box}(\vec{p})$, where the last step is licensed by $M \models \varphi[s] \Leftrightarrow M^* \models \varphi^+[s] \Leftrightarrow M^*, s \models \psi \Leftrightarrow M, s \models \psi^\times$. ∎

Also it is now easy to say what makes $\sim$ special. Note that standard partial formulas are not in general semantically persistent [23], but possess the weaker quality of *external persistence*: if $M = \langle S, R, V\rangle$, $M' = \langle S, R, V'\rangle$, and for all $p \in Prop, s \in S : V(p, s) \sqsubseteq V'(p, s)$ (shortened $M, s \sqsubseteq M', s$; i.e. M' extends the valuation of M), then persistence holds 'pointwise':

$$M, s \models \varphi \Rightarrow M', s \models \varphi \qquad M, s \dashv \varphi \Rightarrow M', s \dashv \varphi$$

This property can be strengthened to another definability result.

Theorem 2.6 *The externally persistent formulas of the extended modal language are precisely those equivalent to $\sim$-free formulas.*

[22] For Z to be a bisimulation the models should have the format $M_i = \langle S_i, \mathsf{R}_i, V_i\rangle$, but the special first order formulation is more convenient here.

[23] I.e. for different situations under different valuations based on the same modal frame; see chapter 4.

Proof sketch: (for the non-trivial side) Let φ be an extended modal formula which is externally persistent for partial models M. Then by proposition 2.4, φ^+ is truth persistent on models M^*. φ^+ can be brought in *negation normal form* φ^*, where each negation immediately precedes an atom p_i (by *double negation, de Morgan's laws,* $\neg\bot = \top$, $\neg\top = \bot$, $\neg\Box = \Diamond\neg$ and $\neg\Diamond = \Box\neg$). Then theorem 2.4 implies that $\mathrm{ST}(\varphi^*)$ is bisimulation invariant and truth persistent with respect to P_i^+, P_i^-, whence a modification of Lyndon's persistence theorem[24] shows that there is a ψ equivalent to $\mathrm{ST}(\varphi^*)$ which is positive in every occurrence of $P_i^\pm$, leaving the restricted quantification unaltered. Whence by retranslation, $(\mathrm{ST}^{-1}(\psi))^\times$ is the desired $\sim$-free formula. The equivalence of φ and $(\mathrm{ST}^{-1}(\psi))^\times$ is then straightforward:
$M \models \varphi[s] \Leftrightarrow M^* \models \varphi^+[s] \Leftrightarrow M^* \models \varphi^*[s] \Leftrightarrow M^* \models \mathrm{ST}(\varphi^*)[s] \Leftrightarrow M^* \models \psi[s] \Leftrightarrow M^* \models \mathrm{ST}^{-1}(\psi)[s] \Leftrightarrow M, s \models (\mathrm{ST}^{-1}(\psi))^\times$. ■

2.5 Conclusion

In this chapter we first tried to accommodate the method of propositional definability by means of semantic conditions to suit the modal language. This structural approach lead to the definition of the *prima facie* paradoxical notion of 'modal truth function'. We obtained a rather gratuitous algebraic characterization and the insight that without direct invocation of the accessibility relation modal definability will be hard to come by in this way: even fairly strong conditions as automorphism invariance did not manage to trigger definability for simple frames.

So we turned to a different, translational perspective. By essentially reducing partial modal logic to classical logic, we were able to provide the more satisfactory result that what makes the translations of modal formulas special among the other first order formulas is their preservation under bisimulations. By similar techniques, we also identified the (externally) persistent modal formulas as being essentially the $\sim$-free formulas.

Apart from adapting the notion 'bisimulation' for partial semantics, we pointed at a number of special bisimulations, such as 'generated submodel', which will be of use further on.

'Wise after the event' we may even speculate whether preservation under bisimulation cannot be used *directly* in the structural approach, thus combining the two views of this chapter. Like generalized quantifiers may be assumed to be invariant under isomorphisms, modal truth functions would have to be preserved under bisimulations. This, however, presupposes the modal truth functions to be indexed for models, and, moreover, a notion of bisimulation restricted to parts of the models. We postpone this interesting but complicated matter to another occasion.

[24] Cf. [Ly59, corollary 2.1] and [Sa91, chapter 5]; yet this point deserves a more elaborate proof, cf. [La88, section 4.3].

Chapter 3

Propositional completeness

As we saw in the introduction to part I, partial logic gives rise to a wealth of possible notions of logical consequence. First we chart the parameters that determine the nature of logical consequence: the underlying notion of validity (for example, 'always true' vs. 'never false') and the notion of rule (for example, relating valid formulas or true formulas). The actual outcome, the set of valid consequences, is influenced by two other factors: the nature of the models (2-, 3- or 4-valued) and, of course, the truth (and falsity) conditions.

Then we investigate into the properties of the resulting notions of consequence. Not all possible combinations lead to different systems and not all systems are equally interesting. Yet some are, and a number of proposals known from the rather diverse literature on partial logic pop up within this framework. Because of their own interest, and to pave the way for the modal extensions in the next chapter, we study the most interesting and natural systems in more detail.

In particular, we provide deductive systems and prove strong completeness with respect to the consequence relation under inspection. To achieve this, the usual Henkin method is adapted to partial logic.

3.1 Introduction

One of the charms of partial logic is its flexibility. Another its richness. This already appeared from the previous chapters on the subject of definability and it also holds for validity: there is no ubiquitous notion of validity in partial logic. In fact there is at least one aspect of the notion of logical consequence that may already turn up in classical logic.

For instance, what do we mean when we say that $\varphi \Rightarrow \psi$ is a valid rule? One possible interpretation is that if φ is a theorem, so is ψ; another that given an (arbitrary) formula φ we can derive ψ. The celebrated rule *Modus Ponens* $(\varphi, \varphi \rightarrow \psi \Rightarrow \psi)$ is used both ways, and in classical propositional logic the difference between these interpretations is fairly small. Perhaps this is the reason that the distinction is often neglected. But already in normal modal logic[1] the difference is important: the rule *Necessitation* ($\varphi \Rightarrow \Box\varphi$) operates on theorems, not on arbitrary formulas. The distinction between what we call *absolute* and *relative* consequence rules is especially relevant for partial semantics, even in the propositional case. Moreover, the distinction presented does not exhaust the possibilities: one might also want to *reduce* the validity of rules to that of formulas. This can be achieved by invoking the deduction theorem as a heuristic principle; then $\varphi \Rightarrow \psi$ is valid iff $\varphi \rightarrow \psi$ is a valid formula.

The choice of the type of rule interacts with the way validity is defined in the semantics. We can choose, among other things, between a 'true everywhere' (*verification*) and 'false nowhere' (*non-falsification* or *falsifiability*) concept of validity. For example, the law of excluded middle $\varphi \vee \neg\varphi$ is valid for coherent situations under the falsification perspective (it is never false), but it is not valid under the verification ('always true') perspective, for the valuation may be undefined with respect to φ. There is yet another option: validity can be obtained by evaluation on a designated subset of the worlds involved in a model. This possibility proves especially relevant for the modal language; one type of this nonstandard approach is treated in chapter 4.

The effect of such distinctions also depends on the sort of indices and valuations used: are they *partial* or *total*, *coherent* or *incoherent*? For example, the relative construal of *Modus Ponens* is verified on three-valued models, but not for four-valued models.

Finally, one may opt for different truth conditions. It was shown in the first chapter that different truth-conditions are compatible with classical logic, for example in the case of implication. But in our general setting even in propositional logic truth conditions with a genuinely modal flavour qualify, in the sense that they make reference to extensions of the situation of interpretation. These 'eventual' conditions are illustrated by intuitionistic logic (section 3.4.4), supervaluation semantics (section 3.4.6) and other systems. Yet in what we consider the *standard* case, truth conditions will not be eventual, but 'actual'. In other words, the standard interpretation of the connectives will be truth-functional. Sections 3.2 and 3.3 deal with completeness in the standard propositional case.

In this chapter and the next one we develop a general framework for truth and validity in partial semantics, controlled by the following parameters.

- situations & valuations: partiality and coherence
- truth-conditions: standard ('actual') vs. non-standard 'eventual'
- validity:(unrestricted) verification (VERIF) or falsifiability (FALSIF) or some restricted set of situations
- type of rule: absolute or relative

In principle no combination of parametrized possibilities is excluded. This may lead to a mixture of the validity type in rules: for example, the approach that reduces consequence to valid implication can be construed as one leading from verified premises

[1] And, similarly, in first order predicate logic.

to a non-falsified conclusion. Generally speaking, quite a number of proposals made in the literature nicely fit into the framework. In particular, it captures almost all truth-functional accounts.[2]. This points at adequacy of the framework and may hopefully lead to more insight in the specific logics, which often seem quite *ad hoc*.

Since the interplay of the different parameters regulating the logic proves to be a complex matter, we have refrained from incorporation of new propositional or modal connectives, and largely stick to the *standard* logical language. So, at least on the syntactic side the logic looks traditional.

3.2 Coherent situation semantics

Presumably the most intuitive form of partial semantics is displayed in the following system, which is usually attributed to [Ba81].[3] Assume the propositional language to be constructed in the obvious way with a set of propositional atoms called *Prop*. Since our prime goal is to model modal rather than mere propositional logic, (and sometimes we even need 'modal' models to interpret propositional languages, cf. section 3.4) the valuation function will be indexed as in section 2.1.[4] Given a model $M = \langle S, V \rangle$, S being a set of coherent situations (or: partial worlds) and V a partial function from $Prop \times S$ to $\{0, 1\}$ (truth values), the usual truth conditions relative to M and $s \in S$ are:[5]

$$\begin{array}{ll} s \models p \Leftrightarrow V(p,s) = 1 \ (\forall p \in Prop) & s \mathrel{=\!\!\!|} p \Leftrightarrow V(p,s) = 0 \ (\forall p \in Prop) \\ s \models \neg\alpha \Leftrightarrow s \mathrel{=\!\!\!|} \alpha & s \mathrel{=\!\!\!|} \neg\alpha \Leftrightarrow s \models \alpha \\ s \models \alpha \wedge \beta \Leftrightarrow s \models \alpha \text{ and } s \models \beta & s \mathrel{=\!\!\!|} \alpha \wedge \beta \Leftrightarrow s \mathrel{=\!\!\!|} \alpha \text{ or } s \mathrel{=\!\!\!|} \beta \end{array}$$

One says that s *verifies* (or *satisfies, supports*) φ whenever $s \models \varphi$ and that s *falsifies* (or *rejects*) φ whenever $s \mathrel{=\!\!\!|} \varphi$. Else, when $s \not\models \varphi$ and $s \not\mathrel{=\!\!\!|} \varphi$, φ is *unknown* or *undefined* in s.

The truth conditions for the other propositional connectives then follow directly from the given conventions and the usual recursive definitions:

- $\alpha \vee \beta := \neg(\neg\alpha \wedge \neg\beta)$,
- $\alpha \rightarrow \beta := \neg\alpha \vee \beta$,
- $\alpha \leftrightarrow \beta := (\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$.

This leads to derived truth conditions for $\vee$, $\rightarrow$ and $\leftrightarrow$:

$$\begin{array}{ll} s \models \alpha \vee \beta \Leftrightarrow s \models \alpha \text{ or } s \models \beta & s \mathrel{=\!\!\!|} \alpha \vee \beta \Leftrightarrow s \mathrel{=\!\!\!|} \alpha \text{ and } s \mathrel{=\!\!\!|} \beta \\ s \models \alpha \rightarrow \beta \Leftrightarrow s \mathrel{=\!\!\!|} \alpha \text{ or } s \models \beta & s \mathrel{=\!\!\!|} \alpha \rightarrow \beta \Leftrightarrow s \models \alpha \text{ and } s \mathrel{=\!\!\!|} \beta \\ s \models \alpha \leftrightarrow \beta \Leftrightarrow s \mathrel{=\!\!\!|} \alpha, \beta \text{ or } s \models \alpha, \beta & s \mathrel{=\!\!\!|} \alpha \leftrightarrow \beta \Leftrightarrow s \mathrel{=\!\!\!|} \alpha, s \models \beta \text{ or } s \models \alpha, s \mathrel{=\!\!\!|} \beta \end{array}$$

[2] An exception is [Bl86] (see section 3.4) which has a different type of validity.

[3] Although the *perspective* of partiality was presumably new, the text clauses were known from multi-valued logic, possibly in a somewhat different format, though for example [Se67] already uses the $\models$, $\mathrel{=\!\!\!|}$ notation.

[4] Likewise, we use the 'modal' notation $M, s \models \varphi$ right from the start.

[5] $M, s \models \varphi$ is abbreviated to $s \models \varphi$ when no confusion arises.

In fact it will be convenient for several results to make the presentation entirely symmetrical by adding $\vee$ as a basic symbol.

Due to the above clauses a situation need not verify classical tautologies, such as the 'law of excluded middle'. So, possibly $s \not\models \varphi \vee \neg\varphi$ (which is intended when employing a truth value gap) and $s \not\models \varphi \rightarrow \varphi$ (which is counterintuitive)[6].

The semantics presented here has the following important properties:

Proposition 3.1 (coherence)
For no $\varphi, S, V, s \in S : M, s \models \varphi$ and $M, s \dashv \varphi$ (where $M = \langle S, V\rangle$).

Proof: by simultaneous induction on the structure of φ. [Th90a, appendix B] ∎

Proposition 3.2 (inherited classicality) [7]
For all $\varphi, M = \langle S, V\rangle, s \in S$ such that V is bivalent ($\mathrm{ran}(V) = 2$):
$M, s \models \varphi \Leftrightarrow M, s \models \varphi \Leftrightarrow M, s \not\dashv \varphi$.

Proof: by simultaneous induction on the structure of φ. ∎

Proposition 3.3 (partiality)
There is a model M and a situation s (called the empty situation) such that for all formulas φ: $M, s \not\models \varphi$ and $M, s \not\dashv \varphi$.

Proof: let $S = \{\frac{1}{2}\}$, and $V(p, \frac{1}{2}) = \frac{1}{2}$ for all $p \in Prop$. The proposition then follows by induction on the structure of φ. ∎

Another property, already familiar from earlier chapters, tells us that more complex facts are known when more basic facts are known, for example when more data have become available. *Persistence* of truth values is defined with respect to the relation of extension. $\sqsubseteq$, which was introduced in chapter 1, is redefined here in the present format. The definition is somewhat more general than usually, because of our intensional preoccupation.

Definition 3.1 ($\sqsubseteq$)
If $s, s' \in S$, $M = \langle S, V\rangle$ and $M' = \langle S, V'\rangle$, then $M, s \sqsubseteq M', s'$ if for every atom p: $V(p, s) = V'(p, s')$ whenever $V(p, s)$ is defined.

Whenever M and M' are clear from the context, and especially when $M = M'$ (*internal* extension), we will write $s \sqsubseteq s'$. If for all $s \in S : M, s \sqsubseteq M', s$, we write $M \sqsubseteq M'$ (*external* extension).

[6] Łukasiewics' proposal 'to fill the gap' by making $\alpha \rightarrow \beta$ true when both antecedent and succedent are undefined (see the appendix to chapter 1) solves this problem but produces many other counterintuitive results, for example when α and β are independent unknown propositions.

[7] Cf. the notion of 'reliability' in [La88, p.18], which is subordinate to classicality + persistence.

Proposition 3.4 (persistence)
If $M, s \sqsubseteq M', s'$, then $M, s \models \varphi \Rightarrow M', s' \models \varphi$, and $M, s \mathrel{=\!\!\!|} \varphi \Rightarrow M', s' \mathrel{=\!\!\!|} \varphi$ for every φ in the standard language.

Proof: by simultaneous induction on the structure of φ. ∎

But the force of a semantics is not solely determined by its truth conditions. One other important factor is validity. In the classical propositional case (i.e. when V is bivalent) a formula is valid iff it is verified everywhere, or, equivalently, falsified nowhere. Interestingly, in a partial setting the notions of verification and falsifiability diverge widely, akin to what comes to mind from the perspective of philosophy of science. We shall discuss both options below.

3.2.1 Propositional verification

Presumably the first possibility which suggests itself is to call a formula valid if it is supported by every situation in each model. So we define

Definition 3.2 (VERIF)
φ is verifiably valid iff for every model $M = \langle S, V \rangle$ and every $s \in S : M, s \models \varphi$ (Notation: $\models \varphi$).

This notion turns out to have weird logical consequences, given the fact that a logic is usually described to a large extent by the set of valid formulas. However, by partiality (proposition 3.3) it follows that

Corollary 3.1 *The set of verifiably valid formulas is empty.*

Since there are no valid formulas according to the verification perspective, there is no point in invoking a deduction theorem which reduces valid consequence to valid implication. So, for verification, the absolute and the reductive approaches to rules are uninformative: these notions would yield the total set of rules ($\mathcal{L} \times \mathcal{L}$) or the empty set ($\emptyset$) respectively. But in the relative approach, verification of full-fledged arguments (i.e. possible rules) is essential since the logic is specified by its *rules* rather than by its *axioms*. What we need is a definition of relative verifiable validity, called strong consequence in [Ba81]:

Definition 3.3 (VERIF$_{rel}$)
$\alpha_1, \ldots, \alpha_m / \beta_1, \ldots, \beta_n$ is relatively verifiably valid iff for every M, s such that for all i: $M, s \models \alpha_i$, for some j: $M, s \models \beta_j$ ($\alpha_1, \ldots, \alpha_m \models \beta_1, \ldots, \beta_n$).

Typically, the rule of *ex falso sequitur quodlibet* is verified, i.e. $\varphi \wedge \neg\varphi \models \psi$. This is an immediate consequence of *coherence*. Its contrapositive, the rule of *tertium non datur*, is not verificationally valid, however: $\psi \Rightarrow \varphi \vee \neg\varphi$ does not hold in general, since for example $p \not\models q \vee \neg q$ (take $V(p, s) = 1, V(q, s) = \frac{1}{2}$). *A fortiori*, the principle of extensionality (substitution under logicaql equivalence) does not hold for this semantics: $\alpha \models\mid \beta$ does not in general imply $\varphi(\alpha) \models\mid \varphi(\beta)$. To wit, $p \wedge \neg p \models\mid p \wedge \neg p \wedge q$ but *not* $\neg(p \wedge \neg p) \models\mid \neg(p \wedge \neg p \wedge q)$.

Notice that *Modus Ponens* is valid for relative verification on coherent models, but does not characterize the set of valid rules on its own. [Ka83] contains such a complete set of inference rules. Though elegant, his characterization invokes a special normal form[8] and corresponding proof technique which does not seem to generalize easily to the modal case. The system we propose here is more direct and closer to natural deduction[9] in classical logic. [10] [11]

(R1) $\neg\neg\varphi \dashv\vdash \varphi$ ('the law of double negation')

(R2) $\neg(\varphi \wedge \psi) \dashv\vdash \neg\varphi \vee \neg\psi$ (first 'de Morgan's law')

(R3) $\neg(\varphi \vee \psi) \dashv\vdash \neg\varphi \wedge \neg\psi$ (second 'de Morgan's law')

(R4) $\varphi \wedge \psi \vdash \varphi \qquad \varphi \wedge \psi \vdash \psi$

(R5) $\varphi \vdash \varphi \vee \psi \qquad \psi \vdash \varphi \vee \psi$

(R6) if $\varphi, \varrho \vdash \chi$ and $\psi, \varrho \vdash \chi$ then $\varphi \vee \psi, \varrho \vdash \chi$

(R7) if $\chi \vdash \varphi, \varrho$ and $\chi \vdash \psi, \varrho$ then $\chi \vdash \varphi \wedge \psi, \varrho$

(R8) $\varphi \wedge \neg\varphi \vdash \psi$ (*ex falso [sequitur quodlibet]*)

(R9) if $\varphi \vdash \psi$ and $\psi \vdash \chi$ then $\varphi \vdash \chi$

(R10) $\Gamma \vdash \Delta$ iff there are nonempty sets $\{\alpha_1, \ldots, \alpha_m\} \subseteq \Gamma$ and $\{\beta_1, \ldots, \beta_n\} \subseteq \Delta$ such that $\alpha_1 \wedge \cdots \wedge \alpha_m \vdash \beta_1 \vee \cdots \vee \beta_n$ $(\alpha_1, \ldots, \alpha_m \vdash \beta_1, \ldots, \beta_n)$

The usual starting rule, viz. $\alpha_1, \ldots, \alpha_m \vdash \beta_1, \ldots, \beta_n$ if some α_i equals some β_j, is provable in this system. It follows from the two observations just below.

Proposition 3.5 (reflexivity) $\varphi \vdash \varphi$

Proof:[12]

1. $\varphi \wedge \varphi \vdash \varphi$ [R4]
2. $\{\varphi\} \vdash \{\varphi\}$ [1, R10]
3. $\varphi \vdash \varphi$ [2, R10]

■

[8] See section 3.4.2.

[9] Note however that we do not adopt Segerberg's terminological distinction between inference rules (such R1–5) and deduction rules (R6,7,9,10), cf. [BS84, p.28].

[10] $\varphi \dashv\vdash \psi$ abbreviates $\varphi \vdash \psi$ & $\psi \vdash \varphi$. The alternative symbols $\Rightarrow$ and $\Leftrightarrow$ are here reserved for meta-level implication and equivalence, respectively.

[11] A concise, yet elegant rule system for the smaller language $\mathcal{L}_{\neg,\wedge}$ is given by [Ur86].

[12] Notice the first application of R10 in the above proof serves to eliminate $\wedge$, the second one to eliminate braces.

Proposition 3.6 (monotonicity)
if $\Gamma \vdash \Delta, \Gamma \subseteq \Gamma'$ *and* $\Delta \subseteq \Delta'$ *then* $\Gamma' \vdash \Delta'$.

Proof: this is an immediate consequence of R10. ■

Other common properties such as commutativity, distributivity and associativity of $\wedge$ and $\vee$, as well as the cut–rule, are derivable.[13]

Proposition 3.7 (commutativity)

- $\varphi \wedge \psi \dashv\vdash \psi \wedge \varphi$
- $\varphi \vee \psi \dashv\vdash \psi \vee \varphi$

Associativity was already used in the formulation of R10, which actually has to be decorated with lots of brackets in order to remain consistent with the order of presentation.

Proposition 3.8 (associativity)

- $\varphi \wedge (\psi \wedge \chi) \dashv\vdash (\varphi \wedge \psi) \wedge \chi$
- $\varphi \vee (\psi \vee \chi) \dashv\vdash (\varphi \vee \psi) \vee \chi$

Proposition 3.9 (distributivity)

- $\varphi \wedge (\psi \vee \chi) \dashv\vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$
- $\varphi \vee (\psi \wedge \chi) \dashv\vdash (\varphi \vee \psi) \wedge (\varphi \vee \chi)$

Proof: We only show the first equivalence:
($\vdash$)

1.	$\varphi \wedge \psi \vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	[R5]
2.	$\varphi, \psi \vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	[1, R10]
3.	$\varphi, \chi \vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	[by analogy]
4.	$\varphi, \psi \vee \chi \vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	[2, 3, R6]
5.	$\varphi \wedge (\psi \vee \chi) \vdash (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$	[4, R10]

($\dashv$)

1.	$\varphi \wedge \psi \vdash \varphi$	[R4]
2.	$\varphi \wedge \psi \vdash \psi$	[R4]
3.	$\psi \vdash \psi \vee \chi$	[R5]
4.	$\varphi \wedge \psi \vdash \psi \vee \chi$	[2, 3, R9]
5.	$\varphi \wedge \psi \vdash \varphi \wedge (\psi \vee \chi)$	[1, 4, R7]
6.	$\varphi \wedge \chi \vdash \varphi \wedge (\psi \vee \chi)$	[by analogy]
7.	$(\varphi \wedge \psi) \vee (\varphi \wedge \chi) \vdash \varphi \wedge (\psi \vee \chi)$	[5, 6, R6]

[13]The proofs are easy exercises, cf. [Th90a, appendix A]. We reprove the first law of distributivity because the proof in [Th90a] was given for the wrong system, viz. Kamp's. The cut–rule is proven because of its importance.

■

Proposition 3.10 (cut)
if $\Sigma \vdash \Delta, \gamma$ *and* $\Sigma, \gamma \vdash \Delta$ *then* $\Sigma \vdash \Delta$.

Proof: Assume $\Sigma \vdash \Delta, \gamma$ and $\Sigma, \gamma \vdash \Delta$ then by R10 there are $\alpha_1, \ldots, \alpha_{k+l} \in \Sigma$ and $\beta_1, \ldots, \beta_{m+n} \in \Delta$ such that $\alpha_1 \wedge \cdots \wedge \alpha_k \vdash \beta_1 \vee \cdots \vee \beta_m \vee \gamma$ and $\alpha_{k+1} \wedge \cdots \wedge \alpha_{k+l} \wedge \gamma \vdash \beta_{m+1} \vee \cdots \vee \beta_{m+n}$ (if γ does not occur in the premise or consequence, we are ready). Let $\alpha = \alpha_1 \wedge \cdots \wedge \alpha_{k+l}$ and $\beta = \beta_1 \vee \cdots \vee \beta_{m+n}$, then by R10: (i) $\alpha \vdash \beta \vee \gamma$ and (ii) $\alpha \wedge \gamma \vdash \beta$. Since (R4,R10) $\alpha \vdash \alpha$, put together with (i) this implies (R7) $\alpha \vdash \alpha \wedge (\beta \vee \gamma)$, and so by the first distribution law and R9: (iii) $\alpha \vdash (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$. By R4, (ii) and R6, $(\alpha \wedge \beta) \vee (\alpha \wedge \gamma) \vdash \beta$, which combined with (iii) by R9 shows $\alpha \vdash \beta$, hence $\Sigma \vdash \Delta$, by R10. ■

Now let $\mathbf{rL}^+$ be the set of deduction rules generated by R1–10. We obtain the following important completeness result.

Theorem 3.1 $\Sigma \models \varphi \Leftrightarrow \Sigma \vdash_{\mathbf{rL}^+} \varphi$.

Proof: It is easily checked that verification on coherent situations is sound for the rules of $\mathbf{rL}^+$. To prove the other direction we use the Henkin-style proof method. So we argue by contraposition. Suppose that $\Sigma \nvdash \varphi$. Now the standard way to extend $\Sigma \cup \{\neg\varphi\}$ does not fit into this semantics, for $\varphi \nvdash \varphi \wedge (\psi \vee \neg\psi)$, yet $\{\varphi, \neg(\varphi \wedge (\psi \vee \neg\psi))\}$ is inconsistent. So we have to proceed more carefully. But the idea is still basically the same: extend Σ to a set Δ such that $\varphi \notin \Delta$ and for which we can prove a truth lemma: $\Delta \models \psi$ iff $\psi \in \Delta$ for each ψ. Then, obviously $\Delta \not\models \varphi$ and we are ready.

Extension of Σ to a suitable Δ is guaranteed by the Lindenbaum lemma below. Such suitable sets of formulas are at least *consistent saturated theories*[14] (CSTs for short), as defined below; they are the partial counterparts of the maximally consistent sets in the classical Henkin proof. Then the situations of the *canonical model* $\mathcal{M} = \langle S, \mathcal{V} \rangle$ are simply the CSTs, and the *canonical valuation* $\mathcal{V}$ is defined by $\mathcal{V}(p, \Gamma) = 1$ iff $p \in \Gamma$ and $\mathcal{V}(p, \Gamma) = 0$ iff $\neg p \in \Gamma$, for all $\Gamma \in S$. Since each Γ is consistent, $\mathcal{V}$ is a well-defined partial function. Then we obtain the truth lemma for all Γ. Since we have $\Sigma \subseteq \Delta$ and $\varphi \notin \Delta$, the truth lemma shows $\mathcal{M}, \Delta \models \Sigma$ and $\mathcal{M}, \Delta \not\models \varphi$, whence $\Sigma \not\models \varphi$. ■

First we define the relevant syntactic notions of *consistency*, *saturation* ('deciding' disjunctions) and *theory* (deductive closure) for an arbitrary set of formulas Σ and an arbitrary inference relation $\vdash$.

- Σ is *consistent* (for $\vdash$) iff $\Sigma \nvdash \varphi \wedge \neg\varphi$ for all φ;
- Σ is *saturated* (for $\vdash$) iff $\Sigma \vdash \varphi$ or $\Sigma \vdash \psi$ for all φ and ψ such that $\Sigma \vdash \varphi \vee \psi$;
- Σ is a *theory* (for $\vdash$) iff $\Sigma \vdash \varphi$ implies $\varphi \in \Sigma$ for all φ.

[14] Cf. 'saturated sets' in [Ac68] and [Th68] and 'CS-theory' in [Ve85].

In other words, if Σ is a theory it is closed under the rules of the deductive system; for reflexive monotone $\vdash$ the converse condition holds, i.e. $\Sigma \vdash \varphi$ iff $\varphi \in \Sigma$. If Σ is a theory then the additional requirement of consistency amounts to $\varphi \wedge \neg\varphi \notin \Sigma$, and saturation to $\varphi \vee \psi \in \Sigma \Rightarrow \varphi \in \Sigma$ *or* $\psi \in \Sigma$. Now let $\vdash$ be the inference relation of $\mathbf{rL}^+$.

Lemma 3.1 (partial Lindenbaum lemma)
If $\Sigma \nvdash \varphi$, then Σ can be extended to a CST *Δ such that $\varphi \notin \Delta$.*

Proof: Let $\Sigma \nvdash \varphi$ and $\varphi_0, \varphi_1, \ldots, \varphi_n \ldots$ be an enumeration of the (well-formed) formulas such that each formula of the language occurs countably many times.[15] Δ_n is defined recursively in such a way that it does not entail φ:

- $\Delta_0 = \Sigma$;
- if $\Delta_{3n} \nvdash \varphi_n$ then $\Delta_{3n+3} = \Delta_{3n+2} = \Delta_{3n+1} = \Delta_{3n}$;
- if $\Delta_{3n} \vdash \varphi_n$ then:
  - $\Delta_{3n+1} = \Delta_{3n} \cup \{\varphi_n\}$;
  - $\Delta_{3n+2} = \Delta_{3n+1} \cup \{\psi\}$ if $\varphi_n = \psi \vee \chi$ and $\Delta_{3n+1}, \psi \nvdash \varphi$, else $\Delta_{3n+2} = \Delta_{3n+1}$;
  - $\Delta_{3n+3} = \Delta_{3n+2} \cup \{\chi\}$ if $\varphi_n = \psi \vee \chi$ and $\Delta_{3n+2}, \chi \nvdash \varphi$, else $\Delta_{3n+3} = \Delta_{3n+2}$.

Let Δ be $\bigcup_n \Delta_n$. Then Δ has the desired properties:

1. Δ is a theory with respect to $\mathbf{rL}^+$: if $\Delta \vdash \psi$ then there are $\delta_1, \ldots, \delta_k \in \Delta$ such that $\delta_1 \wedge \cdots \wedge \delta_k \vdash \psi$. So there is an ℓ with $\delta_1, \ldots, \delta_k \in \Delta_\ell$ and, by the way we defined the enumeration of φ_n, there is an $n \geq \frac{\ell}{3}$ for which $\psi = \varphi_n$. Therefore $\Delta_{3n} \vdash \varphi_n$, and consequently $\psi \in \Delta_{3n+1} \subseteq \Delta$.
2. Δ extends Σ, since $\Sigma = \Delta_0 \subseteq \bigcup_n \Delta_n = \Delta$.
3. Because Δ is a theory, $\varphi \notin \Delta$ if $\Delta \nvdash \varphi$, which in its turn is implied by $\Delta_k \nvdash \varphi$ for every k; this can be shown by induction on k:
   - For $k = 0$ this is given ($\Sigma \nvdash \varphi$).

     Next assume $\Delta_{3n} \nvdash \varphi$, the induction hypothesis (IH)
   - If $k = 3n + 1$, suppose $\Delta_{3n} \vdash \varphi_n$ (the other case is trivial). So $\Delta_{3n+1} = \Delta_{3n} \cup \{\varphi_n\}$. Now suppose $\Delta_{3n+1} \vdash \varphi$. The cut theorem then shows $\Delta_{3n} \vdash \varphi$, which contradicts IH, so $\Delta_{3n+1} \nvdash \varphi$.
   - For $k = 3n+2$ and $k = 3n+3$ the proposition follows directly from the definition of Δ_k.
4. Due to R8, $\Delta \nvdash \varphi$ entails that Δ is consistent with respect to $\mathbf{rL}^+$.
5. Δ is also *saturated*. We will give an indirect proof: assume that $\psi \vee \chi \in \Delta$, yet $\psi \notin \Delta$ and $\chi \notin \Delta$. Thus for some n: $\Delta_{3n} \vdash \psi \vee \chi$, $\Delta_{3n}, \psi \vdash \varphi$, and $\Delta_{3n}, \chi \vdash \varphi$. Then (R6) $\Delta_{3n}, \psi \vee \chi \vdash \varphi$, and so by the cut theorem $\Delta \vdash \varphi$, which contradicts $\varphi \notin \Delta$.

[15]The countable repetition of formulas fascilitates some steps of the proof.An enumeration $\psi_0, \psi_1, \psi_2, \psi_3, \ldots$ of $\mathcal{L}$ can be turned into an enumeration with countable repetition by taking ever larger initial parts: $\psi_0, |\psi_0, \psi_1, |\psi_0, \psi_1, \psi_2, |\psi_0, \psi_1, \psi_2, \psi_3, | \ldots$, which amounts to a sequence $\{\varphi_n\}_n$ where $\varphi_{\frac{1}{2}\ell(\ell+1)+k} = \psi_k$ if $k \leq \ell$.

■

Lemma 3.2 (partial truth lemma)
Let Γ *be a* CST, *then:* $\mathcal{M}, \Gamma \models \psi$ *iff* $\psi \in \Gamma$, *and* $\mathcal{M}, \Gamma \rightmodels \psi$ *iff* $\neg\psi \in \Gamma$ *for all* ψ.

Proof: by simultaneous induction on the structure of ψ:

- (basic case) if ψ is a propositional atom, the lemma holds by the definition of $\mathcal{V}$.

  For the next cases assume the lemma to hold for ψ up to certain complexity, (the induction hypothesis, IH); we will make excessive use of the fact that Γ is a theory.

- let ψ be of the form $\neg\chi$.
  $\Gamma \models \neg\chi$ iff $\Gamma \rightmodels \chi$ iff (IH) $\neg\chi \in \Gamma$.
  $\Gamma \rightmodels \neg\chi$ iff $\Gamma \models \chi$ iff (IH) $\chi \in \Gamma$ iff (R1) $\neg\neg\chi \in \Gamma$.

- let $\psi = \alpha \wedge \beta$.
  $\Gamma \models \alpha \wedge \beta$ iff $\Gamma \models \alpha$ & $\Gamma \models \beta$ iff (IH) $\alpha \in \Gamma$ & $\beta \in \Gamma$ iff (R4 and R7) $\alpha \wedge \beta \in \Gamma$.
  $\Gamma \rightmodels \alpha \wedge \beta$ iff $\Gamma \rightmodels \alpha$ or $\Gamma \rightmodels \beta$ iff (IH) $\neg\alpha \in \Gamma$ or $\neg\beta \in \Gamma$ iff (R5 and saturation) $\neg\alpha \vee \neg\beta \in \Gamma$ iff (R2) $\neg(\alpha \wedge \beta) \in \Gamma$.

- let $\psi = \alpha \vee \beta$.
  $\Gamma \models \alpha \vee \beta$ iff $\Gamma \models \alpha$ or $\Gamma \models \beta$ iff (IH) $\alpha \in \Gamma$ or $\beta \in \Gamma$ iff (R5,saturation) $\alpha \vee \beta \in \Gamma$.
  $\Gamma \rightmodels \alpha \vee \beta$ iff $\Gamma \rightmodels \alpha$ & $\Gamma \rightmodels \beta$ iff (IH) $\neg\alpha \in \Gamma$ & $\neg\beta \in \Gamma$ iff (R4,R7) $\neg\alpha \wedge \neg\beta \in \Gamma$ iff (R3) $\neg(\alpha \vee \beta) \in \Gamma$. ■

Completeness is a very beneficial property, because it allows a shift of perspective (from inference to consequence, or *vice versa*). So, to show that something is derivable, simply check whether it is a valid consequence. But also, the consequence test, which needs a vast class of models, can be replaced by the much more restricted combinatorics of inference. For example, efficient theorem provers based on partial logic are feasible, even though the truth tables are larger than in the classical case. By the property of partiality we can even show the consistency of the deductive system itself (setting $\Sigma = \emptyset$). Assuming a somewhat more general definition of logical consequence ($\Sigma \models \varphi$ in which Σ may be infinite), completeness (non-vacuously) implies *compactness*, which says that a set of formulas is satisfiable iff all its finite subsets are) by moving to the syntactic side. Going the other way, it is now easy to show decidability of the inferential system. What we have not established (and cannot establish by completeness) is the independence of the rules R1–10. Although we want our *characterization* of $\mathbf{rL}^+$ to be as small as possible, this is usually considered a minor issue.

In all, strong consequence and its syntactic counterpart $\mathbf{rL}^+$ are interesting. Yet, the verification approach may still seem strange since it produces no tautologies. So let us pay attention to the other option of the validity type.

3.2.2 Propositional falsifiability

A prime source of motivation of partiality and falsifiability are Beth's semantic tableaus. In a Beth tableau one tests the validity of an inference by trying to construct a counterexample; the (in)validity only depends on propositional variables occurring in the premises and conclusion and often not even all of them! As noticed by [vB84a] the attempted falsification naturally leads to partiality, and, we might add, to falsifiable validity. This perspective can also be found in early work on multi–valued logic. Then sometimes 1 and $\frac{1}{2}$ are the designated truth–values triggering validity, in other words, 0 (and 2 in the 4–valued case) exclude tautologies.

Definition 3.4 (FALSIF)
φ is falsifiably valid ($\not\dashv \varphi$) iff for no model $M = \langle S, V\rangle$ and no $s \in S$: $M, s \mathrel{=\!\!|} \varphi$

For example, $\varphi \vee \neg\varphi$ is valid under this definition simply because it is never rejected. Because valid formulas abound now, we do not really need a separate definition of *consequence*: the notion *falsifiably valid rule* can be reduced to that for *formulas* by stipulating, for example, $\alpha, \beta \not\dashv \gamma \Leftrightarrow \not\dashv (\alpha \wedge \beta) \to \gamma$, i.e. one can employ the familiar deduction theorem not as a derived theorem but as a guiding principle. Since $s \not\dashv (\alpha \wedge \beta) \to \gamma \Leftrightarrow$ if $s \models \alpha$ and $s \models \beta$ then $s \not\dashv \gamma$, the effect is that non–falsity of the conclusion 'mixes' with the truth of the premises. We shall therefore call this type of validity *mixed falsifiability*.

Definition 3.5 (FALSIF$_{mix}$)
$\varphi_1, \ldots, \varphi_n/\psi$ is mixed falsifiably valid iff for every M and s:
if $M, s \models \varphi_1, \ \ldots, \ M, s \models \varphi_n$ then $M, s \not\dashv \psi$.

This notion gives rise to a remarkable 'unpartial' result.

Theorem 3.2 (van Benthem)
FALSIF$_{mix}$ *on coherent models is completely described by classical propositional logic* **pL**.[16]

We can obtain a similar result for *absolute* falsifiable validity.

Definition 3.6 (FALSIF$_{abs}$)
$\varphi_1, \ldots, \varphi_n/\psi$ is absolutely falsifiably valid iff $\not\dashv \varphi_1$ and ...and $\not\dashv \varphi_n$ jointly imply $\not\dashv \psi$.

Absolute rules such as $\vdash \varphi \Rightarrow \vdash \psi$ suffer from a technical complication not yet dealt with. The problem is that in addition to relatively or mixedly valid propositional rules there is a class of absolute rules that qualify for the simple reason that the premise is a contingent formula: for example $\vdash p \Rightarrow \vdash q$ whereas of course $p \not\vdash q$. [17] Without claiming elegance, we can give a very simple solution: let the rules of **pL** be combined with scheme

[16] [vB84a] uses Beth tableaus (or Gentzen sequents) and the notions of strong and weak consequence, where we (would) use ordinary models, and relative verifiable and mixed falsifiable validity, respectively.

[17] This complication seems to have been widely overlooked, but [Cu63, p.97/8,175/6] is very accurate on this point.

$$\vdash \varphi \Rightarrow \vdash \psi \text{ if } \vdash \chi \not\Rightarrow \vdash \varphi,$$

together forming the absolute propositional logic $\mathbf{pL_a}$. But are not we overdoing things? For notice that the above scheme can be reformulated as:

$$\vdash \varphi \Rightarrow \vdash \psi \text{ if } \not\vdash \varphi$$

and does not this follow from the very meaning of the inference relation $\Rightarrow$? The answer to the latter question is that although the 'implication interpretation' is clearly intended, it does not exist *a priori*, for it can only be obtained once a correspondence between deductive system and semantics has been established ... and for the latter we need the above clause! In other words, without completeness the symbols $\vdash$ and $\Rightarrow$ in inference rules are strictly speaking meaningless — they could as well stand for 'is a sentence structure' and 'derived by grammar', respectively. Still, it may be a bit surprising that some common propositional properties are destroyed; for example, contraposition does not hold anymore: in the extended system $p \Rightarrow p \wedge \neg p$, but $p \vee \neg p \not\Rightarrow \neg p$. Notice however that this deviation is not caused by either partiality or coherence; precisely the same observation applies to a classical semantics.[18]

Theorem 3.3 *The coherent semantics with absolute falsifiable validity is complete with respect to the (absolute) propositional logic* $\mathbf{pL_a}$.

Proof: Since our construal of absolute deduction rules with respect to falsifiable validity mimics ordinary logical consequence, the proof reduces to showing that FALSIF valid formulas are the classical tautologies, i.e. $\not\dashv \varphi \Leftrightarrow \models \varphi$ for all φ.
($\Rightarrow$) we argue by contraposition. Suppose that $M, s \not\models \varphi$, then *classical inheritance* shows that M may be considered a coherent partial model, and $M, s \dashv \varphi$. ($\Leftarrow$) again by contraposition: suppose that $M = \langle S, V \rangle$ is a coherent partial model such that $M, s \dashv \varphi$ for some $s \in S$. Then M can be extended to a classical model M^+ by putting, for example, $V^+(p, s) = 1$ iff $V(p, s) \neq 0$ and $V^+(p, s) = 0$ iff $V(p, s) = 0$ (M^+ is called the positive completion of M). Then by *persistence* (proposition 3.4) $M^+, s \dashv \psi$, so by *classical inheritance* $M^+, s \not\models \varphi$. ∎

Now how about *relative* falsifiability, still for coherent partial models?

Definition 3.7 (FALSIF$_{rel}$)
$\varphi_1, \ldots, \varphi_n / \psi$ *is relatively falsifiably valid* ($\varphi_1, \ldots, \varphi_n \not\dashv \psi$) *iff* $M, s \not\dashv \varphi_1$ *and ... and* $M, s \not\dashv \varphi_n$ *jointly imply* $M, s \not\dashv \psi$ *for every* M, s.

Notice that relative verification and falsification are related by contraposition, regardless of the kind of models:

[18] An alternative solution to this problem is to change the notion of 'absolute rule' by imposing the extensionality principle on it. So, in [FHV90], one way (called schematic inference) of construing $\varphi \Rightarrow \psi$ is to require the absolute validities to be related for arbitrary substitutions σ: if $\models \sigma(\varphi)$ then $\models \sigma(\psi)$. Then clearly $p \not\Rightarrow q$ since under one substitution: $\models p \vee \neg p$ but $\not\models q$.

Proposition 3.11 $\varphi \not\dashv \psi$ *iff* $\neg\psi \models \neg\varphi$. [19]

Proof: $\varphi \not\dashv \psi \Leftrightarrow$ for every M, s: if $M, s \not\dashv \varphi$ then $M, s \not\dashv \psi \Leftrightarrow$ for every M, s: if $M, s \dashv \psi$ then $M, s \dashv \varphi \Leftrightarrow$ for every M, s: if $M, s \models \neg\psi$ then $M, s \models \neg\varphi \Leftrightarrow \neg\psi \models \neg\varphi$ ■

Relative falsifiability demonstrates a peculiar behaviour. As we have seen, the set of valid formulas coincides with the set of classical tautologies. However, this does not hold for relative consequence. Crucially, the *ex falso* principle is now invalid. For let p be undefined in s and q be false, then $s \not\dashv p \wedge \neg p$ while $s \dashv q$. Thus the standard (sometimes the *only*) rule in axiomatizations of **pL**, viz. *modus ponens* does not hold either: *ex falso* and *modus ponens* are equivalent, *modulo* the other rules of $\mathbf{rL}^+$. This indicates that valid *formulas* and valid *rules* are independent devices, where a rule that correctly produces tautologies may not qualify as a rule *of* (i.e. within) the same system! By proposition 3.11 FALSIF_{rel} valid rules turn out to be contrapositive to the relatively verified ones. So put

(R8*) $\psi \vdash \varphi \vee \neg\varphi$ (*tertium non datur*),

and let $\mathbf{rL}^*$ be $\mathbf{rL}^+$ with R8 replaced by **(R8*)**.

Theorem 3.4
On coherent models the set of relatively falsifiably valid proper rules is completely described by the system $\mathbf{rL}^*$, *i.e.* $\Sigma \not\dashv \varphi \Leftrightarrow \Sigma \vdash_{\mathbf{rL}^*} \varphi$, *for all* φ *and all* $\Sigma \neq \emptyset$.

The proof of this theorem will be postponed to section 3.3 since it uses a reduction technique active in a more general setting.

In connection to corollary 3.1, we are confronted with another striking result: coherent partial semantics does not select an interesting subset from the set of classical tautologies; it either yields the empty set or else the total set of tautologies. So, although this type of model seems to be better motivated for reasons of intuition and efficiency, at least for valid formulas the outcome is not much different from classical logic. It may therefore have sense to relax the restriction to coherence.

3.3 General situation semantics

Without the restriction to coherence, situations may be incoherent with respect to a proposition and a valuation. We will call the resulting structure a *(general) situation model*. Formally, $V : Prop \times S \longrightarrow \wp(\{0,1\})$, i.e. the valuation function is multiple-valued. The truth and falsity relations are defined in the same way as in section 3.2, with one minor proviso for the basic case:

[19]Notice that this fact is independent of the kind of model employed, as long as the truth conditions for negation stay the same. Moreover, the proposition can be easily generalized for consequence between arbitrary sets.

$$s \models p \Leftrightarrow 1 \in V(p, s), \text{ and } s \mathrel{=\!\!|} p \Leftrightarrow 0 \in V(p, s) \quad (\text{where } p \in Prop).$$

The definition of *coherent situation* has to be modified accordingly. In addition the notion of *total situation* is relevant in the present setting.

Definition 3.8 *(for given $M = \langle S, V\rangle$ and Prop:)*

(totality) *$s \in S$ is* **total** *in M iff for all $p \in Prop : V(p, s) \neq \emptyset$ (else, s is* **partial***);*

(coherence) *s is* **coherent** *in M iff for all $p \in Prop$: $V(p, s) \neq 2 = \{0, 1\}$ (else,* **incoherent***).*

Both properties inductively generalize to arbitrary standard formulas:

Proposition 3.12 (totality)
If s is total in M then for all φ : $M, s \models \varphi$ or $M, s \mathrel{=\!\!|} \varphi$.

Proposition 3.13 (local coherence)
If s is coherent then for all φ : $s \not\models \varphi$ or $s \not\mathrel{=\!\!|} \varphi$.

What kind of logic do general models yield for both definitions of validity?

First, for verification, corollary 3.1 still holds. Some of the relative rules of $\mathbf{rL}^+$, however, are now illegitimate. For example, *Modus Ponens* is invalid: if $s \models p$ and $s \models p \rightarrow q$, then possibly $s \not\models q$, notably when $V(p, s) = \{0, 1\}$ and $1 \notin V(q, s)$. One easily obtains the same result for *Modus Tollens*: $\neg q, p \rightarrow q \not\models \neg p$. There are still valid rules, however, ranging from trivial ones such as $\varphi \models \varphi$ to less trivial such as $\neg\neg\varphi \mathrel{|\!\!=\!\!|} \varphi$ and $\varphi \models \varphi \vee \psi$.

Next for falsifiable validity, we are confronted with a result similar to corollary 3.1, with dual proof: consider a singleton model and a valuation which is overdefined (both true and false) for every atom. Then the model falsifies every formula. So, in general, we obtain for both sorts of validity:

Theorem 3.5 *There are no valid formulas in general situation semantics.*

Notice that proposition 3.11 still holds for general situation models. Instead of enlarging the interrelation of contraposition and validity concepts, however, we can now establish a more revealing connection between the two notions of validity. First we recast the truth-functional notion of *duality* (cf. chapter 1) into a transformation of models.

Definition 3.9 (duality)
For any model $M = \langle S, V\rangle$, its dual $\tilde{M} = \langle S, \tilde{V}\rangle$ is defined by: $1 \in \tilde{V}(p, s)$ iff $0 \notin V(p, s)$, and $0 \in \tilde{V}(p, s)$ iff $1 \notin V(p, s)$.

The effect of this transformation generalizes to complex formulas, since the standard language was shown to be duality preserving in chapter 1.

Proposition 3.14 (duality)
For all M, s, φ : $M, s \not\models \varphi \Leftrightarrow \tilde{M}, s \Dashv \varphi$, and $M, s \not\Dashv \varphi \Leftrightarrow \tilde{M}, s \models \varphi$.

We have paved the way for a useful reduction. The following proposition means that the validity concepts coincide sort by sort: falsifiability and verification of formulas, absolute verification and absolute falsifiability of rules, etcetera.

Proposition 3.15 *For general situation models:* $\not\Dashv \; = \; \models$.

Proof: we shall restrict ourselves to relative consequence, in fact to one side of the simple case $\varphi \Rightarrow \psi$, from which the more general case $\Gamma \models \Delta$ easily follows. So let $\varphi \models \psi$ and $M, s \not\Dashv \varphi$, then (proposition 3.14) $\tilde{M}, s \models \varphi$, hence $\tilde{M}, s \models \psi$ and therefore $M, s \not\Dashv \psi$. Thus $\varphi \not\Dashv \psi$. The other cases are similar. ■

Propositions 3.11 and 3.15 jointly imply that (strong) consequence on general situations is closed under contraposition, and so in this respect the logic is more 'classical' than with coherent situations.

What is the syntactic counterpart of general situation semantics and what is the relation between this rule system and that for verification on coherent models? As we saw before, these two systems are surely different, for example with respect to *Modus Ponens*. Now if we inspect the rules of $\mathbf{rL}^+$, *ex falso* is typically not valid on general models, but the other rules are. Indeed, the set of rules $\mathbf{rL}^+ - \{R8\}$ (called **rL** henceforth) proves to be complete with respect to relative validity on general situations.

Theorem 3.6 *The system* **rL** *is complete for general consequence.*

Proof: Soundness of **rL** is easily checked. The completeness of the rule system, i.e. that $\Gamma \models \varphi$ implies $\Gamma \vdash_{\mathbf{rL}} \varphi$ is shown by 'armchair–reasoning'. Just reinspect the proof for theorem 3.1. Use saturated theories STs instead of CSTs. Notice that R8 is not used in the proof or its lemmas, apart from the check on coherence of Δ in the Lindenbaum lemma. But this step is now superfluous. Apart from one minor technical detail this indeed completes the proof. Since the valuation may be overdefined we must define $\mathcal{V}$ by: $1 \in \mathcal{V}(p, \Delta)$ iff $p \in \Delta$, and $0 \in \mathcal{V}(p, \Delta)$ iff $\neg p \in \Delta$. ■

This rule system coincides with one already existing. One way to reveal this is to translate general situation models into the 4-valued total ones encountered before. The 'underlying' sets of truth-values $\emptyset, \{0\}, \{1\}, \{0, 1\}$ are identified with the new values $\frac{1}{2}, 0, 1$ and 2, respectively. So, the new value 2 is used if p was overdefined in s, and $\frac{1}{2}$ when p is underdefined. Likewise, the truth/falsity conditions given at the beginning of this chapter correspond to the truth tables in chapter 1. These tables also pop up in relevance logic, see [Be77]. In fact what can be shown is that the logic of general situations provides the rules for Belnap's relevant propositional logic! (... and that is why we dubbed the logic **rL**.) Instead of giving a syntactic proof of the equivalence, we shall point out how the semantics of both systems are related. Recall the partial order $\leq$ of the four truth-values as displayed in the logical lattice (from bottom to top):

$$\begin{tikzcd}[column sep=small] & 1 \arrow[dl, dash] \arrow[dr, dash] & \\ \frac{1}{2} \arrow[dr, dash] & & 2 \arrow[dl, dash] \\ & 0 & \end{tikzcd}$$

('Tautological') entailment, i.e. relevant consequence, can be defined by means of the logical ordering. φ entails ψ (notated here as $\varphi \leq \psi$) holds iff for every model $M = \langle S, V\rangle$ and every $s \in S : [\![\varphi]\!](s) \leq [\![\psi]\!](s)$ where $[\![\cdot]\!]$ extends V to interpretation of arbitrary formulas.

Now if we compare $\leq$ to verifiable consequence, at first sight the two notions seem to diverge. If $[\![\varphi]\!](s)$ and $[\![\psi]\!](s)$ have the values $\frac{1}{2}$ and 0 (or: $\frac{1}{2}$ and 2, or: 1 and 2) respectively, then s is no counterexample for $\varphi \models \psi$, but *is* a counterexample for $\varphi \leq \psi$. What is at stake here, is that we are comparing the consequence relations *locally*, on just one model and one situation, although they express *universal* facts: in order to establish consequence, we have to consider *all* models. From this global perspective we can argue as follows: assume one of the three problematic cases to occur, for example the first one. Then although s is not a counterexample to $\varphi \models \psi$, it is one for $\neg\psi \models \neg\varphi$, so by propositions 3.11 and 3.15, $\varphi \models \psi$ cannot hold. In the other cases, and in the other direction, we argue analogously. So, $\models \;\Leftrightarrow\; \leq$.

An advantage of the more general setting is that it permits useful proof techniques. One illustration of this was the reduction of partial logic to 'duplicated' classical logic executed in chapter 2.

Also the general setting allows for important subclasses: one is the set of coherent models, another the set of total models. These subclasses are related by the duality transformation. Reconsidering proposition 3.15 we may note that duality transforms non-falsifying coherent models into verifying total ones. Equipped with one extra syntactic notion we are ready to prove theorem 3.4.

- Σ is *full* (for $\vdash$) iff $\Sigma \vdash \varphi \vee \neg\varphi$ for all φ.

So, non-empty **rL***–theories are full. Notice moreover that, modulo saturation, the full sets are exactly the deductively complete sets[20] ($\Sigma \vdash \varphi$ or $\Sigma \vdash \neg\varphi$ for all φ.)

Theorem 3.4 (repeated)
On coherent models, $\Sigma \not\dashv\!\!\!\!\dashv \varphi \Leftrightarrow \Sigma \vdash_{\mathbf{rL}^*} \varphi$ for all φ and all $\Sigma \neq \emptyset$.

Proof: Soundness of **rL*** with respect to FALSIF$_{rel}$ is straightforward. For completeness, the considerations given above show that FALSIF$_{rel}$ on coherent situations amounts to VERIF$_{rel}$ on total ones. To prove that VERIF$_{rel}$ on total models is complete with respect to **rL*** reconsider the proof of theorem 3.1. As already shown for **rL** the truth lemma holds without reference to R8 (or to R8*). Since the Δ constructed in the Lindenbaum lemma contains Σ and is closed

[20] Also called 'complete theories', which is extremely confusing in the present context since it uses different meanings of 'completeness' and 'theory' (viz. arbitrary set of formulas) than used here.

under **rL***, it is a full saturated theory (FST). For each FST Γ we have $p \vee \neg p \in \Gamma$, so (by saturation) $p \in \Gamma$ or $\neg p \in \Gamma$, which implies $1 \in \mathcal{V}(p, \Gamma)$ or $0 \in \mathcal{V}(p, \Gamma)$, thus $\mathcal{V}(p, \Gamma) \neq \emptyset$. Hence Γ is total. The rest of the proof remains the same. ∎

3.4 Alternatives in partial semantics

Before we turn to 'rival' proposals, let us make a few remarks about validity and completeness for the extended propositional language. As noticed at the end of section 3.1 we refrained from treating these aspects of the expanded language *in extenso*, since the effectuation of the announced program turned out to be a fairly elaborate task, even when restricting to the basic propositional language.

3.4.1 extending the language

In chapter 1 two non-standard negations were considered: $\sim$ and ∂. Each of them has advantages of its own. $\sim$ can be interpreted both in 3- and in 4-valued semantics. ∂ can only be interpreted properly in 4-valued models. So, $\sim$ is more flexible in this respect. We repeat the crucial truth tables of the standard negation $\neg$ and the two non-standard negations below.

	1	$\frac{1}{2}$	0	2
$\neg$	0	$\frac{1}{2}$	1	2

	1	$\frac{1}{2}$	0	2
$\sim$	0	1	1	0

	1	$\frac{1}{2}$	0	2
∂	0	2	1	$\frac{1}{2}$

These tables show that, unlike $\sim$, the semantic operation corresponding to ∂ is self-inverse, and in this respect ∂ behaves as $\neg$. The negations $\sim$ and ∂ can also be contrasted with $\neg$ in its truth and falsity clauses.

$$\begin{array}{lll} s \models \neg\varphi \Leftrightarrow s \mathrel{=\!\!|} \varphi & s \models \sim\varphi \Leftrightarrow s \not\models \varphi & s \models \partial\varphi \Leftrightarrow s \not\models \varphi \\ s \mathrel{=\!\!|} \neg\varphi \Leftrightarrow s \models \varphi & s \mathrel{=\!\!|} \sim\varphi \Leftrightarrow s \models \varphi & s \mathrel{=\!\!|} \partial\varphi \Leftrightarrow s \not\mathrel{=\!\!|} \varphi \end{array}$$

So, $\neg$ is the (symmetrical) standard negation, ∂ a symmetrical nonstandard negation, and $\sim$ an asymmetrical nonstandard negation, which mixes the truth and falsity conditions of $\neg$ and ∂ in one way.[21]

It can easily be ascertained that ordinary tautologies such as *tertium non datur* are verified w.r.t. non-standard negation, i.e. $\models \varphi \vee \sim\varphi$ and $\models \varphi \vee \partial\varphi$. In other words, there is a very classical ring to $\sim$ and ∂.

By means of these new[22] negations, we can define non-standard implications $\supset$ and $\rightharpoonup$ in the usual way. We include the standard implication for comparison.

$$\varphi \to \psi = \neg\varphi \vee \psi \qquad \varphi \supset \psi = \sim\varphi \vee \psi \qquad \varphi \rightharpoonup \psi = \partial\varphi \vee \psi$$

[21] The other possible mixture is described by $\neg \sim \neg$.

[22] It follows from results in chapter 1 that $\sim$ and ∂ are not definable by means of standard connectives, even when supplied with constants: neither is *persistent*. Also $\sim$ and ∂ are not interdefinable, modulo the standard connectives: ∂ is not *generally closed* (as $\neg, \wedge, \sim$ are), and $\sim$ is not *duality preserving* (but $\neg, \wedge, \partial$ are).

This leads to the following truth tables for these implications:

$\rightarrow$	1	$\frac{1}{2}$	0	2
1	1	$\frac{1}{2}$	0	2
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1
0	1	1	1	1
2	1	1	2	2

$\supset$	1	$\frac{1}{2}$	0	2
1	1	$\frac{1}{2}$	0	2
$\frac{1}{2}$	1	1	1	1
0	1	1	1	1
2	1	$\frac{1}{2}$	0	2

$\rightharpoonup$	1	$\frac{1}{2}$	0	2
1	1	$\frac{1}{2}$	0	2
$\frac{1}{2}$	1	1	2	2
0	1	1	1	1
2	1	$\frac{1}{2}$	$\frac{1}{2}$	1

Despite their differences, $\supset$ and $\rightharpoonup$ have an important feature in common: with respect to verification they both formalize the notion of *strong consequence*. The reason for this is that they share the partial truth condition, which is, in some sense, more natural than the one for $\rightarrow$.

$$s \models \varphi \supset \psi \Leftrightarrow s \models \varphi \rightharpoonup \psi \Leftrightarrow \ s \models \varphi \Rightarrow s \models \psi$$

In other words, it is the falsity clause that distinguishes $\supset$ and $\rightharpoonup$.

Now what are the tautologics obtained in the extended languages? We do have classical tautologies again, for example, $\models \varphi \supset \varphi$ and $\models \varphi \supset (\psi \supset \varphi)$. Since the latter is one of the usual axiom schemes for classical propositional logic, we might be led to think that the classical tautologies in $\mathcal{L}_{\neg,\supset}$ are precisely the partially valid ones. This would be blatantly wrong, for at least two reasons. First, not all classical tautologies in this language are strongly valid, for example $(\neg q \supset \neg p) \supset (p \supset q)$ is *not* partially valid (choose $V(p, s) = 1$ and $V(q, s) = \frac{1}{2}$. Second, not all partial validities could be obtained in this way, for example, $\models \varphi \supset (\psi \supset (\varphi \wedge \psi))$, but $\wedge$ cannot be expressed in $\mathcal{L}_{\neg,\supset}$.

The other extended language, with ∂ instead of $\sim$ is similar, to a large extent. The main reason for this is the fact that $\sim$and ∂ are (verificationally) equivalent:

$$\sim \varphi \models\!\!\!\mid \partial\varphi$$

Again there are the usual validities such as $\models \varphi \rightharpoonup (\psi \rightharpoonup \varphi)$ and $\models \varphi \rightharpoonup (\psi \rightharpoonup (\varphi \wedge \psi))$, whereas $(\neg q \rightharpoonup \neg p) \rightharpoonup (p \rightharpoonup q)$ is not valid.[23]

Notice, however, that there are some differences between these two extensions, which are caused by the interaction of standard and non-standard negations. In particular, we have $\neg\partial\varphi \not\models \neg \sim \varphi$ and $\neg \sim \varphi \not\models \neg\partial\varphi$. One striking difference between $\sim$ and ∂ is that $\neg \sim \varphi \models\!\!\!\mid \varphi$, but *not* $\neg\partial\varphi \models\!\!\!\mid \varphi$. Another that $\neg\partial\varphi \models\!\!\!\mid \partial\neg\varphi$, but *not* $\neg \sim \varphi \models\!\!\!\mid \ \sim \neg\varphi$. We believe that addition of a number of such principles to the standard system (for 3- or 4-valued strong consequence) yields a complete deductive system. We postpone the details of such a complete description to another occasion.

Another interesting remark concerns notions of validity. In general the greater expressive force of the extended language makes it possible to restrict validity to verification: for example, relative falsifiable validity may now be formulated as $\sim \neg\varphi_1, \ldots, \sim \neg\varphi_n \models \ \sim \neg\psi$.

[23] So, $\rightharpoonup$ resembles the conditional occurring in the relevance logic **R**, cf. [Du86].

The standard systems of 'pure partial logic' studied in this chapter can be applied to and compared with other proposals.[24]

3.4.2 Kamp

Following the original work of [Ba81] and [BP81], [Ka83] is treating the situation semantics of perception verbs with formal rigour. Kamps's truth conditions for the connectives and his notion of validity are standard-type. So, disregarding the ontology, the semantics as such is standard.

The propositional part of the deductive system presented in [Ka83] deviates from our system $\mathbf{rL}^+$: R7 is replaced by two rules:

(R0) if $\varphi \Rightarrow \psi$ and χ is a positive context of φ, then $\chi(\varphi) \Rightarrow \chi(\psi)$

(R7′) $\varphi, \psi \Rightarrow \varphi \wedge \psi$

R0 contains a simple but useful concept: if $\varphi, \chi \in \mathcal{L}_{\neg,\wedge,\vee}$, then χ is said to be a *positive context* for φ, if φ is a subformula in χ occurring in a positive position, i.e. not within the scope of a negation. The application of this rule is more widespread than might appear at first sight. A formula is then in *negation normal form* (NNF) if its negations occur only in front of *atoms*, i.e. in literals.[25] For example, $\neg(\neg p \vee q)$ is equivalent to $p \wedge \neg q$. So by the above method:

Lemma 3.3 *Every formula is verifiably equivalent to a formula in* NNF.

Let $\mathbf{rL}'$ be the set of deduction rules generated by R0–10, with R7 replaced by R7′. After showing the cut rule we may reprove completeness of $\mathbf{rL}'$ with respect to relative verification.

Consequently, $\mathbf{rL}$ and $\mathbf{rL}'$ are equivalent. This can, of course, also be shown by a direct syntactic argument.[26]

In all, we conclude that Kamp's approach is not essentially different from ours. The alternatives to follow are really non-standard. We shall start with the approach of [Bl86].

3.4.3 Blamey

[Bl86] deals (mainly) with coherent situations and the usual connectives with standard interpretation, but in addition there are some 'funny connectives', as he puts it: apart from the constants $\star$, $\top$ and $\bot$, one encounters *interjunction* and *transplication*, here symbolized by $\bowtie$ and $\hookrightarrow$:[27]

---

[24] [Th90a] treats these proposals more extensively; a short discussion of [Mu89] is included at the end of chapter 4.

[25] The NNF of φ can also be obtained by back and forth translation: $(\varphi^+)^\times$, cf. chapter 2. See chapter 4 for the more specific *disjunctive* normal form.

[26] For example, R7′ follows from R4 and R10: by *reflexivity*, $\varphi \wedge \psi \vdash \varphi \wedge \psi$, and R10 implies $\{\varphi, \psi\} \vdash \{\varphi \wedge \psi\}$. R0 can be derived by induction on the structure of the positive context.

[27] [Mu89] notices that $\bowtie$ amounts to the *meet* operation $\sqcap$ on the approximation (semi-)lattice.

$\times\!\!\times$	0	$\frac{1}{2}$	1
0	0	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	$\frac{1}{2}$	1

$\hookrightarrow$	0	$\frac{1}{2}$	1
0	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
1	0	$\frac{1}{2}$	1

So $\times\!\!\times$ is a very weak counterpart of both $\wedge$ and $\vee$, whereas $\hookrightarrow$ is a weak implication. It follows from chapter 1 that neither is definable in terms of the standard connectives. Blamey notices that $\times\!\!\times$, $\hookrightarrow$ and $\star$ are interdefinable, modulo the other connectives. Each one added to the standard connectives $\neg$, $\wedge$, and $\top$ provides definability of persistent truth functions.

Where $\star$ and $\times\!\!\times$ are fairly artificial, $\hookrightarrow$ is motivated by its application to presupposition phenomena. If ψ is the presupposition of φ and χ the assertoric contents of φ, then φ may be construed as $\psi \hookrightarrow \chi$, which accounts for the *negation test*[28]: if φ is true and if $\neg\varphi$ is true, ψ must be true. So we may now consistently require $\varphi \mathrel{|\!\!\equiv} \psi$ and $\neg\varphi \mathrel{|\!\!\equiv} \psi$ without running into inconsistencies for contingent ψ.

Although the construal with $\hookrightarrow$ is a major step forward in the slippy field of presupposition, phenomena such as presupposition cancellation are still problematic. For example, (with φ, ψ, χ as above) the sentence 'If ψ then φ' does not have the presupposition ψ, and, roughly, means $\psi \rightarrow \chi$. But $\psi \hookrightarrow (\psi \hookrightarrow \chi)$ is strongly equivalent to φ, and the other possible translation $\psi \rightarrow (\psi \hookrightarrow \chi)$ is also not equivalent to $\psi \rightarrow \chi$.

With respect to logical consequence and equivalence Blamey argues for a so-called "double-barrelled" approach, combining relative verification and falsification:[29]

$$\begin{array}{lll} \varphi \models \psi & \text{iff} & \varphi \mathrel{|\!\!\equiv} \psi \;\&\; \varphi \not\mathrel{\equiv\!\!|} \psi \\ \varphi \mathrel{|\!\!=\!\!|} \psi & \text{iff} & \varphi \models \psi \;\&\; \psi \models \varphi \end{array}$$

Blamey notices that $\mathrel{|\!\!\equiv}$ and $\not\mathrel{\equiv\!\!|}$ can be defined in terms of (double-barrelled) $\models$, using $\star$:

$$\begin{array}{lll} \varphi \mathrel{|\!\!\equiv} \psi & \text{iff} & \varphi \models \psi \vee \star \\ \varphi \not\mathrel{\equiv\!\!|} \psi & \text{iff} & \varphi \wedge \star \models \psi \end{array}$$

For the standard language we notice that double-barrelled consequence on general situations is axiomatized by **rL**; this follows from proposition 3.15. On coherent situations double-barrelled consequence is characterized by $\mathbf{rL}^{+*}$, which is $\mathbf{rL}^{+}$ with R8 replaced by the rule $\mathrm{R8}^{+*}$:

($\mathbf{R8}^{+*}$) $\varphi \wedge \neg\varphi \vdash \psi \vee \neg\psi$

$\mathbf{rL}^{+*}$ is slightly stronger than **rL**, though weaker than $\mathbf{rL}^{+}$ and $\mathbf{rL}^{*}$. [30]

[28]The negation test says that ψ is a presupposition of φ if both φ and *not*–φ logically imply ψ.

[29]Blamey uses $\simeq$ where we use $\mathrel{|\!\!=\!\!|}$, and $\models^{\top}$ and $\models^{\perp}$ for our $\mathrel{|\!\!\equiv}$ and $\not\mathrel{\equiv\!\!|}$; notice that in this subsection $\models$ differs from the usual two-valued notion.

[30]Cf. the 'mixed' system in [Ve87]; though technically correct, it seems intuitively dubious to accept $\mathrm{R8}^{+*}$ on the one hand, and reject the transparant principles R8 and R8* on the other.

Though, once we have one of the extra connectives, Blamey's concept of validity and our two basic concepts of validity are interdefinable, he holds that the double-barrelled approach is supported by "arguments stemming at least from theoretical neatness".[31] An advantage of $\models\!|$ is that it triggers strict identity of truth functions. However, the redefinitions of the verification and non-falsification are quite artificial. By contrast, double-barrelled consequence can easily be derived from our standard notions of validity.

Notice that the approaches of Kamp and Blamey are strictly truth–functional. A different direction employed in several proposals is to 'modalize' the truth conditions by means of the relation $\sqsubseteq$, usually restricted to one model at the time. We will discuss a few of these proposals below. Apart from the truth conditions and some *ad hoc* features, the following types of semantics are standard with respect to the other parameters. For example, validity may be chosen to be verification.

3.4.4 Intuitionistic logic

A well-known theory with a clear intensional flavour is intuitionistic logic. As such the semantical approaches of Beth and Kripke to intuitionistic logic are obviously not partial. Still it is said that intuitionism is to be understood partially: the constructive method urges partiality.

Following Gödel's reduction of propositional intuitionistic logic to a fragment of the modal system **S4**, assigning a 'hidden box' interpretation to atoms, negation and implication, Kripke proposed a possible world semantics to this effect. The evident problem is thus: can we give an implementation of Beth's and Kripke's ideas within the partial framework? Actually the solution to this problem is already indicated by one of Kripke's remarks. [Kr65b] notices that $V(p, w) = 0$ should not be read as 'p has been proved *false* at w', but as 'p has not (yet) been proved, verified'. We will implement this idea as follows: let V be a partial function of atom-world pairs to values in $\{1\}$, i.e. a *partial 1-valued function*. Notice that no Kripkean stipulation for persistence (with respect to alternatives) is needed, once we replace the relation of accessibility by that of extension. So what is really different then from standard accounts of partiality is in the truth conditions, which now become:

$$
\begin{aligned}
&s \models p \Leftrightarrow V(p, s) = 1\\
&s \models \neg\varphi \Leftrightarrow \forall s' \sqsupseteq s : s' \not\models \varphi\\
&s \models \varphi \wedge \psi \Leftrightarrow s \models \varphi \;\&\; s \models \psi\\
&s \models \varphi \vee \psi \Leftrightarrow s \models \varphi \text{ or } s \models \psi\\
&s \models \varphi \to \psi \Leftrightarrow \forall s' \sqsupseteq s : s' \models \varphi \Rightarrow s' \models \psi
\end{aligned}
$$

[31] [Bl86, p.6,7]. However, (i) unlike Blamey we do not believe contraposition to be a necessary ingredient of a logical system: this is supported by intuitions concerning incomplete knowledge, cf. chapter 9; (ii) by definition, logical equivalence is mutual consequence and (iii) notice that $\varphi \dashv\vdash \varphi \wedge \psi \Leftrightarrow \psi \dashv\vdash \psi \vee \varphi \Leftrightarrow \varphi \vdash \psi$ also holds for our systems $\mathbf{rL}^+$ and $\mathbf{rL}^*$.

So the semantics is not bivalent in the classical sense, but may still be considered two–valued (or rather $1\frac{1}{2}$–valued!).

3.4.5 Humberstone

Viewed from our present perspective [Hu81] treats partiality in a deviant, asymmetric way. Situations (*possibilities* in Humberstone's terminology) are indeed partial and coherent with respect to propositional variables. Yet there is but one single truth relation $\models$, and a non-classical, seemingly intuitionistic clause for negation:

$$s \models \neg\varphi \Leftrightarrow \text{ for all } s' \sqsupseteq s : s' \not\models \varphi$$

The basic clause and the truth condition for conjunction are as usual (that is to say, with $\models\!\!\!\equiv$ replaced by $\models$). Introducing the other connectives by the usual definitions, this leads to rather complex and unintuitive truth conditions for $\vee$ and $\rightarrow$. Validity is of the *relative* type; of course the difference between verification and falsification does not play a rôle here.

To illustrate this semantics we notice that, for arbitrary propositions φ and situations s, it may be the case that neither $s \models \varphi$ nor $s \models \neg\varphi$. In *this* sense the models are indeed partial. However, the effect of partiality in complex formulas is immediately reduced (for example, $\varphi \vee \neg\varphi$ is always verified) and, in fact, the logic produced is entirely classical. This can only be achieved by including $\sqsubseteq$ in the overt model structure and impose constraints on admissible frames: *persistence* and *refinability*[32] have to hold for the model structures. In our opinion, reliability is a bit strange a condition on frames. In fact we can easily envisage models where part of the formulas remain forever undecided. Apart from this, the semantics is remarkably complex given that it characterizes classical propositional logic.

3.4.6 van Fraassen

While Humberstone used partial worlds to establish a total interpretation, we encounter the opposite situation in van Fraassens's *supervaluation* semantics, where total worlds serve to derive a partial interpretation. So, let $\langle S, V\rangle$ be a fixed coherent model, let w range over for possible worlds (i.e. coherent and total situations), reserving s for arbitrary coherent situations, then the 'supertruth' conditions are:

$$\begin{aligned} s \models\!\!\!\equiv \varphi &\Leftrightarrow \forall w \sqsupseteq s : w \models \varphi \\ s =\!\!\!\mid \varphi &\Leftrightarrow \forall w \sqsupseteq s : w \not\models \varphi \end{aligned}$$

Notice that the partiality of V is only used in defining the relation $\sqsubseteq$, not in defining any relation of satisfaction. Incidentally, the restriction to extensions of s is absent in [vF66], but this condition is surely in the spirit of the original theory — in fact van Fraassen has formulated more general accounts of supervaluation.

[32] See chapter 1.

The supervaluation approach has some advantages, especially with regards to *penumbral* truths, as [Fi75a] calls it. Suppose in some situation s we do not have information whether p is true or false. Then, intuitively, the formulas p, $p \vee p$ and $\neg p$ are not true, nor are they false. But independent of this indeterminacy, we feel that $p \vee \neg p$ is true: in the 'real world' p has to be either true or false, and in both cases $p \vee \neg p$ is true. In fact all classical tautologies are valid in the supervaluation semantics.

We conclude that the supervaluation semantics is interesting and has certain advantages for describing phenomena such as vagueness, but is not very flexible: formally spoken, it is merely a non-standard semantics for classical logic. Yet [Bl86] seems to us too harsh in his judgement: the fact that partial truth is not persistent under supervaluations should not be surprising since it is not truth–functional either, and the latter intensionality is very much intentional. But can we really blame intensional semantics being intensional?

3.4.7 Veltman

Still in the spirit of intuitionism, but now in an overtly partial fashion is Frank Veltman's *data semantics*, discussed in [Ve81] and [Ve85]. The basic semantic entities are called '(possible) information states', corresponding to our coherent situations, now with the obvious intention to represent correct but possibly incomplete information. The format of an 'information model' is $\langle S, \sqsubseteq, V \rangle$, where V is persistent with respect to the partial order $\sqsubseteq$ among situations. Again $\sqsubseteq$ is included in the model structure to set a constraint on it, which roughly corresponds to Humberstone's refinability[33]:

> **the Zorn property**
> Every maximal chain (linearly ordered subset) of situations contains a maximal element; moreover, such maximal elements are total. [34]

The validity type is that of *relative verification.* The truth and falsity conditions for atoms, $\neg$, $\vee$ and $\wedge$ are standard-type. With regards to the other clauses data semantics went through a number of transitions during its development.

Implications are treated in a somewhat intuitionistic way; in [Ve81] the conditions are:

$$s \models \varphi \rightarrow \psi \text{ iff } \forall s' \sqsupseteq s : s' \models \varphi \Rightarrow s' \models \psi$$
$$s \mathrel{=\!\!|} \varphi \rightarrow \psi \text{ iff } \exists s' \sqsupseteq s : s' \models \varphi \;\&\; s' \mathrel{=\!\!|} \psi$$

Notice the definition is still intrinsically partial: the truth value of $\varphi \rightarrow \psi$ may be undefined in a situation s. This partial effect has been eliminated in [Ve85], where the truth condition is changed into:

$$s \models \varphi \rightarrow \psi \text{ iff } \forall s' \sqsupseteq s : s' \models \varphi \Rightarrow s' \not\mathrel{=\!\!|} \psi$$

[33] Or, rather, CCLOS as a condition on (possibly infinite) sets of situations.

[34] 'Maximal' means 'impossible to extend properly', for chains w.r.t. $\subseteq$, for situations w.r.t. $\sqsubseteq$. The baptizing of the constraint is ours and reminiscent of one of the early equivalents of the Axiom of Choice in axiomatic set theory, i.e. Zorn's lemma.

The modal flavour of these conditions is also explicitly present: the logical language has some modal operators, with a semantics defined by means of the $\sqsubseteq$ relation. Veltman's *must*-operator is interpreted by means of $\sqsubseteq$ as accessibility relation:

$$s \models \Box\varphi \text{ iff } \forall s' \sqsupseteq s : s' \not\models \varphi$$
$$s \models\!\!| \Box\varphi \text{ iff } \exists s' \sqsupseteq s : s' \models\!\!| \varphi$$

The operator $\Diamond$ (*may*) is dual to $\Box$, which already yields its truth conditions. Again the interpretation of modal formulas turns out to be *total*, quite different from what would expect for a partial semantics. However, Veltman gives detailed motivation for his truth conditions, and in fact the whole theory provides a convincing account of the behaviour of conditional sentences.

With regards to the variations in the semantics we have not yet been complete. A different route was suggested in [Ve81][35]:

$$s \models \Box\varphi \text{ iff } \forall \textit{ total } w \sqsupseteq s : w \models \varphi$$
$$s \models\!\!| \Box\varphi \text{ iff } \exists \textit{ total } w \sqsupseteq s : w \models\!\!| \varphi$$

To complicate matters further, the form of data semantics presented in [vB84a] is a 'cross-section' of the previous options: $\rightarrow$ is treated 'partially', $\Box$ is given the above 'supervaluation' account, and $\Diamond$ the earlier 'total' clauses. So in all we are left with four different forms of data semantics. Without going in great detail here we notice that the deductive properties of these logics are also different. For example, the argument $\varphi \rightarrow \psi, \Box\varphi \Rightarrow \Box\psi$ holds in [Ve81] and [vB84a], but not in [Ve85], and the duality principle $\Box\varphi \Rightarrow \neg\Diamond\neg\varphi$ is not validated in [vB84a], but is in the other systems.

An important result is that similar to earlier findings for intuitionistic logic, data logic as discussed in [vB84a] can be reduced to (a subset of) the ordinary modal system **S4.1**.[36]

[Ve85] gives a complete recursive definition of the set of valid rules; in fact his completeness proof inspired early attempts leading to some results in section 4.3 here.

3.5 Conclusion

Different values of semantic parameters (such as validity, kind of model, type of rule) resulted in various systems of logic. The relation between semantics and deductive systems has been given in terms of completeness theorems. These results are summarized in table 3.1. In order to be fully systematic, we have supplemented the picture by adding some results that were not stated in the main text, but follow easily from it.

First, notice we have achieved symmetry in the table by fully exploring *total situations* (which may be overdefined!). The completeness of **rL*** for relative verification

[35] *Vide* [Ve81], footnote 15.

[36] Presumably the eclectic nature of van Benthem's version serves to show that no matter which of the suggested truth conditions are chosen, a translation of data logic into normal modal logic can always be given — [Ve85, p.207] already gives an adapted translation, and it is an easy exercise to accomodate the translation for the actual variants given in [Ve81].

was already demonstrated in the proof of theorem 3.4. By means of the duality operation (section 3.3), validity on total models can be reduced to validity on coherent ones, meanwhile switching from verification falsification.

Second, we discussed *mixed falsifiable validity* on coherent models as one of the alternative ways of characterizing classical propositional logic **pL**. How about mixed *verifiable* validity on coherent models? In other words, which inference rules φ/ψ have the property that $s \not\models \varphi$ implies $s \models \psi$? Notice this is equivalent to requiring $s \models \varphi \rightarrow \psi$, i.e. to the validity of implications. But no standard formula is verifiably valid on coherent models, so the set of rules searched for is empty.

Third, there is one possibility which has been disregarded so far: nothing prohibits *absolute* rules with mixed validity type. In this way one can conclude from 'always verified' to 'never falsified', or the other way round. It turns out, however, that this move only provides one new system of a rather pathological nature, where the premises of the conclusion are non-tautologies and the conclusion arbitrary. This system manifests itself for mixed absolute verification on coherent models: validity of the inference φ/ψ amounts to $\not\models\!\!\!\!\!\dashv \varphi \Rightarrow\ \models \psi$, thus to $\not\vdash_{\mathbf{pL}} \varphi$. These strange rules also pop up in a complete characterization of absolute rules with 'straight' validity type: the resulting system $\mathbf{pL_a}$ is the union of $\mathbf{pL}$ and $\mathbf{pL}^c \times \mathcal{L}$. The default case for absolute and relative rules is still the one in which validity is straight; mixed is the exception.

The different forms of validity and consequence are summatized below:

VERIF $\models \varphi$ iff for all $M, s : M, s \models \varphi$.

VERIF$_{abs}$ $\varphi \Rightarrow \psi$ iff *if* $\models \varphi$ *then* $\models \psi$.

VERIF$_{rel}$ $\varphi \Rightarrow \psi$ iff for all M, s: *if* $M, s \models \varphi$ *then* $M, s \models \psi$. $(\varphi \models \psi)$

VERIF$_{abs,mix}$ $\varphi \Rightarrow \psi$ iff *if* $\not\dashv \varphi$ *then* $\models \psi$.

VERIF$_{mix}$ $\varphi \Rightarrow \psi$ iff for all M, s: *if* $M, s \not\dashv \varphi$ *then* $M, s \models \psi$.

FALSIF $\not\dashv \varphi$ iff for all $M, s : M, s \not\dashv \varphi$.

FALSIF$_{abs}$ $\varphi \Rightarrow \psi$ iff *if* $\not\dashv \varphi$ *then* $\not\dashv \psi$.

FALSIF$_{rel}$ $\varphi \Rightarrow \psi$ iff for all M, s: *if* $M, s \not\dashv \varphi$ *then* $M, s \not\dashv \psi$. $(\varphi \not\dashv \psi)$

FALSIF$_{abs,mix}$ $\varphi \Rightarrow \psi$ iff *if* $\models \varphi$ *then* $\not\dashv \psi$.

FALSIF$_{mix}$ $\varphi \Rightarrow \psi$ iff for all M, s: *if* $M, s \models \varphi$ *then* $M, s \not\dashv \psi$.

We chart our main completeness results in the table 3.1.

Especially relative validity turns out to be interesting, both in theory and in application.[37] How are the various systems of relative consequence related to each other? Of course restricting the type of situations leads to an extension of the set of

[37] See [Ta92] for a fruitful combination of partial logic ($\mathbf{rL}^+$ and $\mathbf{rL}$) and non-monotonic reasoning.

Table 3.1: partial propositional logics

	possible worlds	coherent situations	total situations	general situations
VERIF	**pL**	$\emptyset$	**pL**	$\emptyset$
VERIF_{rel}	**pL**	$\mathbf{rL}^+$	$\mathbf{rL}^*$	**rL**
VERIF_{mix}	**pL**	$\emptyset$	**pL**	$\emptyset$
FALSIF	**pL**	**pL**	$\emptyset$	$\emptyset$
FALSIF_{rel}	**pL**	$\mathbf{rL}^*$	$\mathbf{rL}^+$	**rL**
FALSIF_{mix}	**pL**	**pL**	$\emptyset$	$\emptyset$

valid rules.[38] This state of affairs can be displayed in the following diagram of twin lattices of types of models and the systems they relatively verify:

$$\begin{tikzcd}[column sep=small] & \mathbf{pL} \arrow[dl, dash] \arrow[dr, dash] & \\ \mathbf{rL}^+ \arrow[rr, leftrightarrow, "contrap."] \arrow[dr, dash] & & \mathbf{rL}^* \arrow[dl, dash] \\ & \mathbf{rL} & \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & \text{coh+tot} \arrow[dl, dash] \arrow[dr, dash] & \\ \text{coh} \arrow[rr, leftrightarrow, "dual."] \arrow[dr, dash] & & \text{tot} \arrow[dl, dash] \\ & \text{gen} & \end{tikzcd}$$

The logical systems are related by inclusion (lines, upwards) and contraposition (arrows); the types of semantics are also connected by inclusion (lines, downwards) and dualization (arrows). Are these structures really lattices? Yes, although perhaps not in a self-evident way. The join of $\mathbf{rL}^+$ and $\mathbf{rL}^*$ indeed describes classical propositional inference, in a somewhat redundant natural deduction style. The real problem resides in the lower half of these structures: the intersection of the systems $\mathbf{rL}^+$ and $\mathbf{rL}^*$ is not the system **rL** since it contains a rule which is not in **rL**, viz. the rule '*ex falso sequitur tertium non datur*' $\varphi \wedge \neg\varphi \vdash \psi \vee \neg\psi$, which we recognize as R8^{+*}, the typical rule of Blamey's system $\mathbf{rL}^{+*}$, which is $\mathbf{rL}^+$ with R8 replaced by R8^{+*}. So, if we want the meet operation to correspond with intersection of full inference systems, the bottom element of the lattice should be $\mathbf{rL}^{+*}$(and **rL** could be added below $\mathbf{rL}^{+*}$). Yet the displayed lattices are correct when not the full systems but their *finite descriptions* (R1–10 and the like) are intended.

In the next chapter the combination of partiality and intensionality will be exploited more intensively, with full-fledged accessibility instead of the relation of extension.

[38] More formally, let $\mathcal{M}$ be a class of models validating the system $\mathbf{S}_{\mathcal{M}}$. If $\mathcal{M} \subseteq \mathcal{N}$ then $\mathbf{S}_{\mathcal{N}} \subseteq \mathbf{S}_{\mathcal{M}}$. For example, if $(\varphi \vdash \psi) \in S_{\mathcal{N}}$, $M \in \mathcal{M}$ and $M, s \models \varphi$, then $M \in \mathcal{N}$ and so $M, s \models \psi$.

Chapter 4

Modal completeness

4.1 Introduction and program

The intensional aspect implicit in various proposals for propositional logics (see section 3.4) can and should be generalized to an approach merging partial and modal logic. This chapter will show that such a generalization is feasible and interesting. This in itself does not provide sufficient motivation for such an intricate enterprise. Yet, although the later chapters will elucidate this further, some contemplation may already make the point.

Notice the earlier intensionality depends on the relation of extension $\sqsubseteq$. We saw that $\sqsubseteq$ may of course be used to define the semantics of the modal operators $\Box$ and $\Diamond$. Though this approach is important and interesting, it is limited by the fact that extension is a fixed relation. One of the good things of modal logic is its flexibility: it may be used for a large number of applications, ranging from logical necessity (the *alethic* interpretation of modals), over knowledge (*epistemic*), belief (*doxastic*), ethics (*deontic*) to computer science (the *dynamic* interpretation). Much of this diversity is semantically controlled by different choices of the accessibility relation in possible world models. Although interpreting modals by plain extension still allows some freedom in truth conditions and validity, this is not enough for the observed diversity.

In all, our approach is in the spirit of ordinary Kripke semantics, with partiality, and possibly incoherence, permitted in the valuation. Apart from the dimension of accessibility, there is considerable liberty in partial semantics, as we saw in previous chapters, both for variable (chapters 1 and 2) and for fixed logical languages (chapter 3).

We have explored this new area and found completeness theorems for the '*minimal*' modal logics corresponding to standard types of partial semantics, without any constraint on accessibility. Despite the preparatory work in the preceding chapter, proving completeness by means of the Henkin method turns out to be a complicated matter.

Some general model theory for partial modal logics was, in fact, already present in section 2.3, providing basic techniques such as *bisimulation*. We believe such a transfer of classical to partial modal logic to be possible on a larger scale, but we confine ourselves to methods that will be used further on in this thesis.

To gain insight in the behaviour of *specific* systems, we discuss another useful technique (*filtration*) and give completeness results for a number of the most obvious systems, restricted to relative validity.

Finally we discuss what we consider a deviation from the main trail, where the accessibility relation itself is partialized (see section 4.5.2).

4.2 Possible worlds revisited

Before we turn to truly partial models we reinspect possible world semantics. Why do this? Surely, verification and non-falsification amount to the same on classical Kripke models. And the set of valid formulas is consequently the same in both perspectives, viz. the simple normal system **K**. This system is axiomatized by, for example:

(pL) the axioms and rules (especially *modus ponens*) of **pL**;

(K) $\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$; [1]

(N) if $\vdash \varphi$ then $\vdash \Box\varphi$.[2]

The usual mode of verification of modal rules is the *absolute* approach. But even then we have to be careful to add the principle of absolute closure introduced in section 3.2; otherwise not all valid rules are derivable from **K**). In the sequel we shall use '$\mathbf{K_a}$' to denote the augmented absolute system, which is the system **K** with the additional rule:

$$\vdash \varphi \Rightarrow \vdash \psi \text{ if } \nvdash \varphi$$

More important is that already for classical worlds the relative approach yields a system of rules different from **K**. The point is that the nature of the rules is essential in modal logic. For **N** we note that $\varphi \Rightarrow \Box\varphi$ is valid when construed as an absolute rule, but not as a relative rule, i.e.

$$\varphi \vdash \Box\varphi$$

does not qualify as a rule of the relative system. So we are confronted with the paradoxical situation that **N** can be involved in deriving valid formulas, though it does not qualify as a valid rule itself (cf. the similar case expressed by theorem 3.4).

But what is the characterizing system for relative rules? Apart from **pL** one needs the following inference rules:

[1] Although the formal language does not contain $\rightarrow$ we can reconstruct $\wedge$ and $\vee$ in terms of $\neg$ and $\rightarrow$ and derive all valid formulas this way; alternatively, one might prefer to replace **K** directly by $\Box\varphi \vee \Diamond\psi \vee \Diamond(\neg\varphi \wedge \neg\psi)$.

[2] **N** may be restricted to *axioms*, cf. *Universal Generalization* in first-order logic.

($\mathbf{I_r}$) if $\varphi \vdash \psi$ then $\Box\varphi \vdash \Box\psi$ [3]

($\mathbf{C_r}$) $\Box\varphi \wedge \Box\psi \vdash \Box(\varphi \wedge \psi)$

However, valid principles such as $p \Rightarrow \Box(p \vee \neg p)$ are not derivable from the combination of **pL**, $\mathbf{I_r}$ and $\mathbf{C_r}$ alone. In fact we have to restore **N**, again construed as a relative rule, but now properly:

($\mathbf{N_r}$) $p \rightarrow p \vdash \Box(p \rightarrow p)$.

$p \rightarrow p$ here serves as an arbitrary tautology; because of $\mathbf{I_r}$, any other classical tautology would do equally well. Now let $\mathbf{K_r}$=$\mathbf{pL} + \mathbf{I_r} + \mathbf{C_r} + \mathbf{N_r}$. To motivate this nomenclature, notice that axiom **K** relativized as $\Box(\varphi \rightarrow \psi) \vdash \Box\varphi \rightarrow \Box\psi$ is derivable from $\mathbf{K_r}$. More formally, the systems **K** and $\mathbf{K_r}$ are related by the *deduction theorem*:

Proposition 4.1 $\varphi \vdash_{\mathbf{K_r}} \psi \Leftrightarrow \vdash_{\mathbf{K}} \varphi \rightarrow \psi$

Proof: by induction on the length of the respective derivation. The basic observation underlying this is that the systems can simulate each other.
($\Rightarrow$) It is well-known that **pL**, **I**, **C** are derivable in **K**, see e.g. [Ch80].[4] Also, an application $\mathbf{N_r}$ can be imitated in **K**, since by **pL** $\vdash p \rightarrow p$, thus by **N** $\vdash \Box(p \rightarrow p)$. and by **pL** again $p \rightarrow p \vdash \Box(p \rightarrow p)$.
($\Leftarrow$) This boils down again to checking axioms and rules.

pL: as above;

N: assume $\vdash \varphi$, then, by **pL**, $p \rightarrow p \vdash \varphi$, thus ($\mathbf{I_r}$) $\Box(p \rightarrow p) \vdash \Box\varphi$ and so, by **pL** and $\mathbf{N_r}$, $\vdash \Box\varphi$;

K: because of **pL** $\varphi \wedge (\varphi \rightarrow \psi) \vdash \psi$, thus, by $\mathbf{I_r}$, $\Box(\varphi \wedge (\varphi \rightarrow \psi)) \vdash \Box\psi$, so, by $\mathbf{C_r}$ and **pL**, $\Box\varphi \wedge \Box(\varphi \rightarrow \psi)) \vdash \Box\psi$, and finally, by **pL** (apply the deduction theorem twice), $\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$. ■

So $\mathbf{K_r}$ is **K** in disguise.

This enables us to formulate a completeness theorem for the set of (relative) consequences on ordinary Kripke models. In classical models the distinction between true and not-false disappears; consequently, VERIF and FALSIF (both in straight and mixed mode) coincide. Then the so-called 'weak' completeness theorem for **K** (characterizing valid *formulas*, see e.g. [Ch80]) implies the 'strong' completeness theorem for valid *consequences* in normal modal logic, because of proposition 4.1.

Theorem 4.1
The possible world semantics with relative validity is complete with respect to system $\mathbf{K_r}$, *i.e.* $\Sigma \vdash_{\mathbf{K_r}} \Delta$ *iff* $\Sigma \models \Delta$.

In all, the resulting logic is pretty much like the old system **K** where absolute rules are circumvented. More changes are to be expected for partial or incoherent models.

[3] $\mathbf{I_r}$ and $\mathbf{C_r}$ are the relativized counterparts of **I**: $\vdash \varphi \rightarrow \psi \Rightarrow \vdash \Box\varphi \rightarrow \Box\psi$ and **C**: $\vdash (\Box\varphi \wedge \Box\psi) \rightarrow \Box(\varphi \wedge \psi)$.

[4] Recall from section 3.5 that the '**pL**' in $\mathbf{K_r}$ will be a system of natural deduction, whereas usually the '**pL**' in **K** is obtained from only 3 axioms and *modus ponens*. Yet, these systems, though very different in appearance, are equivalent.

4.3 Coherent modal models

Especially for epistemic applications coherent modal models are of interest: we cannot have inconsistent knowledge.

On the technical level, partial model theory for modal logics combines partial propositional semantics and the possible worlds approach to modalities. Recall from the introduction to part I and chapter 2 that a coherent modal model (or: a partial Kripke model) is a Kripke frame with a partial valuation. More precisely, a partial Kripke model is a triple $\langle S, R, V\rangle$, where $R \subseteq S \times S$ is an accessibility relation and V is a partial function into $\{0, 1\}$.

The truth conditions for the connectives are as stated in section 3.2. In addition the most plausible conditions for the modal operators $\Box$ and $\Diamond$ given earlier in chapter 2 are:

$$\begin{array}{ll} M, s \models \Box\varphi \Leftrightarrow \forall t \in R[s] : M, t \models \varphi & M, s \mathrel{=\!\!\!|} \Box\varphi \Leftrightarrow \exists t \in R[s] : M, t \mathrel{=\!\!\!|} \varphi \\ M, s \models \Diamond\varphi \Leftrightarrow \exists t \in R[s] : M, t \models \varphi & M, s \mathrel{=\!\!\!|} \Diamond\varphi \Leftrightarrow \forall t \in R[s] : M, t \mathrel{=\!\!\!|} \varphi \end{array}$$

The full standard language will contain both $\Box$ and $\Diamond$, but the above clauses allow redefining $\Diamond$ as $\neg\Box\neg$, which is convenient for inductive proofs.

Without explicitly stating or (inductively) proving all of them, we notice that most basic properties of coherent propositional models, viz. *coherence*, *partiality* and *inherited classicality* hold in the modal case *mutatis mutandis*. This is illustrated by the following proposition.

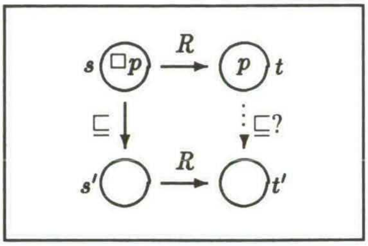
Proposition 4.2 (partiality)
There is a model M and a situation s such that for all formulas φ: $M, s \not\models \varphi$ and $M, s \not\mathrel{=\!\!\!|} \varphi$.

Proof: let $S = \{\frac{1}{2}\}$, $\frac{1}{2}R\frac{1}{2}$ and $V(p, \frac{1}{2}) = \frac{1}{2}$ for all $p \in Prop$. The proposition then follows by induction on the structure of φ. ■

By contrast, general[5] persistence for modal formulas may be violated. The intuitive reason for this violation is that persistence in its general form has a *local* character (comparing single situations), whereas the modal truth clauses have a more *global* nature, involving accessible situations.

To be more specific, imagine a model in which $s \models \Box p$, $s \sqsubseteq s'$ and suppose there is only one t such that sRt. So $t \models p$. Then for any t' which is R-accessible from s' it should be the case that $t' \models p$, but nothing urges $t \sqsubseteq t'$. For there is no compelling reason why the relations in the following diagram should commute.

[5] In particular: *internal* persistence does not hold, cf. section 3.2.



On the basis of this consideration a concrete counterexample for general persistence is easily constructed (take, for example $S = \{s, s', t, t'\}$, $R = \{\langle s, t\rangle, \langle s', t'\rangle\}$, $V(p, t) = 1$ and $V(p, t') = 0$).

Fortunately, a weaker form of persistence does hold. Suppose we merely extend the valuation, in other words employ what we called an *external extension* of the model. Recall (from page 66, adapted for the modal case) that $M \sqsubseteq M'$ if $M = \langle S, R, V\rangle$, $M' = \langle S, R, V'\rangle$, and for every $s \in S : M, s \sqsubseteq M', s$. Then persistence holds 'pointwise' with respect to this special type of extension.

Proposition 4.3 (external persistence)
If $M \sqsubseteq M'$ then $M, s \models \varphi \Rightarrow M', s \models \varphi$ **and** *$M, s \dashv \varphi \Rightarrow M', s \dashv \varphi$ for all standard formulas φ.*

Proof: Due to general persistence for the standard propositional language we only have to check the steps for the modal operators in an inductive proof. Assume the lemma for some φ and all $s \in S$ (IH). Let $M, s \models \Box\varphi$ and $M \sqsubseteq M'$. Then for every t such that sRt: $M, t \models \varphi$, and so by IH $M', t \models \varphi$, and therefore $M', s \models \Box\varphi$. That $M, s \dashv \Box\varphi$ implies $M', s \dashv \Box\varphi$ is shown analogously. ■

Like in the propositional case, for coherent models there is no general account of the validated rules: the distinction between verification and non-falsification is important again.

4.3.1 Modal verification

As in the propositional case, the absolute and mixed approaches are hardly interesting for verification on (partial) worlds: the empty set of validities induces, for example, the total or empty set of rules. Much of this is implied by the observation that for coherent models there are no verifiably valid standard formulas. For it follows by partiality (proposition 4.2) that

Corollary 4.1 *The set of verifiably valid standard modal formulas is empty.*

So we focus on *rules* in the relative perspective. Fortunately the relative system is interesting and certainly non-trivial in its deductive performance. With regard to the problem of completeness, it is clear that the rule system should contain $\mathbf{rL}^+$ (*vide* section 3.2), and a number of characteristic modal rules (R11–19 below), together forming the system $\mathbf{M}^+$, which is the modal counterpart of the logic $\mathbf{rL}^+$ for strong consequence.

(R11) $\Diamond\neg\varphi \dashv\vdash \neg\Box\varphi$

(R12) $\Box\neg\varphi \dashv\vdash \neg\Diamond\varphi$

(R13) $\Box\varphi \wedge \Box\psi \vdash \Box(\varphi \wedge \psi)$ (called $\mathbf{C_r}$ before)

(R14) $\Diamond(\varphi \vee \psi) \vdash \Diamond\varphi \vee \Diamond\psi$

(R15) if $\varphi \vdash \psi$ then $\Box\varphi \vdash \Box\psi$ (called $\mathbf{I_r}$ before)

(R16) if $\varphi \vdash \psi$ then $\Diamond\varphi \vdash \Diamond\psi$

(R17) $\Box\varphi \wedge \Diamond\psi \vdash \Diamond(\varphi \wedge \psi)$

(R18) $\Box(\varphi \vee \psi) \vdash \Diamond\varphi \vee \Box\psi$

(R19) $\Diamond(\varphi \wedge \neg\varphi) \vdash \psi$ (*modal ex falso*).

Some comments may help to clarify aspects of this deductive system:

1. Notice that most rules of $\mathbf{M}^+$ come in dual pairs; except for the 'ex falso' principles R8 and R19, all displayed rules are subject to *contraposition*, in the sense that if $\varphi \vdash \psi$ then $\neg\psi \vdash \neg\varphi$.

2. We conjecture that the rules of the system are independent. The claim is based on abortive attempts to reduce the system, as well as on the fact that dual pairs are not interreducible since contraposition does not hold in general. Anyway, independence is of minor importance compared with consistency and completeness.

3. The rules obtained when disjunction and conjunction in R13 and R14 are interchanged, as well as the converses of R13 and R14, also qualify (the derivations use R15 and R16). We shall demonstrate the latter:

1	$\varphi \vdash \varphi \vee \psi$	[R5]
2	$\Diamond\varphi \vdash \Diamond(\varphi \vee \psi)$	[1, R16]
3	$\Diamond\psi \vdash \Diamond(\varphi \vee \psi)$	[by analogy]
4	$\Diamond\varphi \vee \Diamond\psi \vdash \Diamond(\varphi \vee \psi)$	[2,3, R6]

4. R18 is equivalent to $\Box(\varphi \rightarrow \psi) \vdash \Box\varphi \rightarrow \Box\psi$, which is a relative counterpart of the **K**-axiom, already encountered in section 4.2

Notice that the important properties of $\vdash$ such as reflexivity, monotonicity and the cut-rule still hold for $\mathbf{M}^+$, since $\mathbf{M}^+$ extends $\mathbf{rL}^+$. Consequently, the partial pendant of the Lindenbaum lemma is also valid. We arrive at a central completeness result.

Theorem 4.2 *On coherent models* VERIF$_{rel}$ *is sound and complete with respect to the modal system* $\mathbf{M}^+$, *i.e.* $\Sigma \models \varphi \Leftrightarrow \Sigma \vdash_{\mathbf{M}^+} \varphi$.

Proof: It is easy to check that verification on coherent models is sound for the rules of $\mathbf{M}^+$. To prove the other direction we use a Henkin-style proof. So we argue by contraposition: suppose that $\Sigma \not\vdash \varphi$. Then by the partial Lindenbaum lemma (lemma 3.1) Σ can be extended to a CST Ω such that $\varphi \notin \Omega$. Define the canonical model $\mathcal{M} = \langle \mathcal{S}, \mathcal{R}, \mathcal{V} \rangle$ by: (especially the 'twofold' definition of $\mathcal{R}$ is crucial)

- $\mathcal{S} = \{\Gamma \mid \Gamma \text{ is a CST}\}$;
- $\Gamma \mathcal{R} \Delta$ iff
  - $\Gamma, \Delta \in \mathcal{S}$,
  - $\Box\psi \in \Gamma$ implies $\psi \in \Delta$ for all ψ, and
  - $\psi \in \Delta$ implies $\Diamond\psi \in \Gamma$ for all ψ;
- $\mathcal{V}(p, \Gamma) = 1$ iff $p \in \Gamma$; $\mathcal{V}(p, \Gamma) = 0$ iff $\neg p \in \Gamma$.

Since $\Sigma \subseteq \Omega$ and $\varphi \notin \Omega$, the truth lemma below shows: $\mathcal{M}, \Gamma \models \psi$ iff $\psi \in \Gamma$ for all $\psi \in \mathcal{L}, \Gamma \in \mathcal{S}$. So, *a fortiori*, $\mathcal{M}, \Omega \models \Sigma$ and $\mathcal{M}, \Omega \not\models \varphi$, whence $\Sigma \not\models \varphi$. ∎

To facilitate the modal steps of the inductive proof of the truth lemma, we need some sublemmas. The complexity of these additional lemmas is presumably caused by the twofold character of $\mathcal{R}$. Yet a simple 'single' definition of canonical accessibility is impossible, because the two defining clauses are independent. Here is a simple counterexample, in which the first clause holds but the second one does not: let $\Gamma = \emptyset$ and Δ be a CST containing p. There is a dual counterexample with $\Delta = \emptyset$ and Γ a CST containing $\Box p$ that shows a similar irreducibility in the other direction. For later applications we also give a reformulation of $\mathcal{R}$ in a more concise format. Since $\Box$ and $\Diamond$ are syntactic operations on $\mathcal{L}$, we may adopt the usual functional notation.[6] So $\Gamma \mathcal{R} \Delta$ amounts to $\Box^{-1}[\Gamma] \subseteq \Delta$ and $\Diamond[\Delta] \subseteq \Gamma$, where the latter conjunct is equivalent to $\Delta \subseteq \Diamond^{-1}[\Gamma]$. So,

Lemma 4.1 $\Gamma \mathcal{R} \Delta \Leftrightarrow \Box^{-1}[\Gamma] \subseteq \Delta \subseteq \Diamond^{-1}[\Gamma]$

The proofs of what are essentially the modal steps of the partial truth lemma are certainly less easy.[7]

Lemma 4.2 $\Diamond\psi \in \Gamma$ *iff for some* Δ *such that* $\Gamma \mathcal{R} \Delta$: $\psi \in \Delta$.

Proof: From the right to the left this follows immediately from the definition of $\mathcal{R}$. To show the other direction assume that for fixed ψ and CST Γ: $\Diamond\psi \in \Gamma$. Let $\varphi_0, \varphi_1, \ldots, \varphi_n \ldots$ be an

---

[6] $f[X] = \{f(x) \mid x \in X\}$ and $f^{-1}[X] = \{x \mid f(x) \in X\}$.

[7] We are indebted to Frank Veltman for his help in setting up these proofs; the 'diamond properties' were suggested by Johan van Benthem. The proofs are quite long and may perhaps be skipped at first reading. Yet we would like to encourage the reader to try them on occasion, since this is in some sense the heart of the matter.

enumeration of the formulas of the modal language such that each formula occurs countably many times. The required CST Δ is defined in such a way that for all ε:

$$(\Diamond 1) \qquad \Delta \vdash \varepsilon \Rightarrow \Diamond\varepsilon \in \Gamma,$$

which is called the *first diamond-property* for Δ. To achieve this, Δ_n is defined recursively:

- $\Delta_0 = \Box^{-1}[\Gamma] \cup \{\psi\}$;
- $\Delta_{n+1} = \Delta_n$ if $\Delta_n \nvdash \varphi_n$;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n\}$ if $\Delta_n \vdash \varphi_n$ and φ_n is not a disjunction;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ if $\Delta_n \vdash \varphi_n$ and $\varphi_n = \chi \vee \chi'$ and for all ε: $\Delta_n, \chi \vdash \varepsilon \Rightarrow \Diamond\varepsilon \in \Gamma$;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi'\}$ if $\Delta_n \vdash \varphi_n$ and $\varphi_n = \chi \vee \chi'$ and not for all ε: $\Delta_n, \chi \vdash \varepsilon \Rightarrow \Diamond\varepsilon \in \Gamma$.

Let Δ be $\bigcup_{n\in\omega} \Delta_n$. We will show $(\Diamond 1)$ for Δ_n, i.e. $\Delta_n \vdash \varepsilon \Rightarrow \Diamond\varepsilon \in \Gamma$ by induction on n.

- Assume $\Delta_0 \vdash \varepsilon$, then by R10 (and R4) there are $\delta_1, \ldots, \delta_m \in \Box^{-1}[\Gamma]$ such that $\delta \wedge \psi \vdash \varepsilon$ (where $\delta = \delta_1 \wedge \cdots \wedge \delta_m$). With R16 and R17 this implies $\Box\delta \wedge \Diamond\psi \vdash \Diamond\varepsilon$ (1). Since $\Box\delta_1, \ldots, \Box\delta_m \in \Gamma$, R13 and the fact that CSTs are closed under conjunction, we obtain $\Box\delta \in \Gamma$. Moreover, by assumption, $\Diamond\psi \in \Gamma$, and so, since Γ is a theory, $\Box\delta \wedge \Diamond\psi \in \Gamma$, hence, using (1), $\Diamond\varepsilon \in \Gamma$.
- Suppose Δ_n has $\Diamond 1$ (IH), and consider Δ_{n+1}.
  - In case $\Delta_n \nvdash \varphi_n$ there is nothing to prove.
  - If $\Delta_n \vdash \varphi_n$, and φ_n is not a disjunction, let $\Delta_{n+1} \vdash \varepsilon$, i.e. $\Delta_n, \varphi_n \vdash \varepsilon$. Then the cut-rule yields $\Delta_n \vdash \varepsilon$ and so by IH: $\Diamond\varepsilon \in \Gamma$.
  - If $\Delta_n \vdash \varphi_n = \chi \vee \chi'$ and for all ε': $\Delta_n, \chi \vdash \varepsilon' \Rightarrow \Diamond\varepsilon' \in \Gamma$, then Δ_{n+1} has $(\Diamond 1)$. For let $\Delta_n, \varphi_n, \chi \vdash \varepsilon$, then (cut) $\Delta_n, \chi \vdash \varepsilon$ and thus by definition $\Diamond\varepsilon \in \Gamma$.
  - The last possibility is that in which $\Delta_n \vdash \varphi_n$, $\varphi_n = \chi \vee \chi'$ and for some ε_1: $\Delta_n, \chi \vdash \varepsilon_1$ and $\Diamond\varepsilon_1 \notin \Gamma$. We will show that $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi'\}$ has $(\Diamond 1)$. For suppose not. Then there exists an ε_2 such that $\Delta_n, \varphi_n, \chi' \vdash \varepsilon_2$ and $\Diamond\varepsilon_2 \notin \Gamma$. Consequently, (by R5 and R6) $\Delta_n, \varphi_n, \chi \vee \chi' \vdash \varepsilon_1 \vee \varepsilon_2$, thus by cut: $\Delta_n \vdash \varepsilon_1 \vee \varepsilon_2$, so (IH) $\Diamond(\varepsilon_1 \vee \varepsilon_2) \in \Gamma$. With R14, $\Diamond\varepsilon_1 \vee \Diamond\varepsilon_2 \in \Gamma$, and by saturation of Γ: $\Diamond\varepsilon_1 \in \Gamma$ or $\Diamond\varepsilon_2 \in \Gamma$, which contradicts the assumptions.

Since Δ_n has the $\Diamond 1$-property, so has Δ itself! The rest of the proof is easy.

1. Δ is a theory, for if $\Delta \vdash \chi$ then there are $\delta_1, \ldots, \delta_k \in \Delta$ such that $\delta_1 \wedge \cdots \wedge \delta_k \vdash \chi$. So there is an ℓ with $\delta_1, \ldots, \delta_k \in \Delta_\ell$ and by the way we defined the enumeration of φ_n: there is an $n \geq \ell$ for which $\chi = \varphi_n$. Therefore $\Delta_n \vdash \varphi_n$, and consequently $\chi \in \Delta_{n+1} \subseteq \Delta$.
2. Δ is consistent, for assume on the contrary that for some χ: $\Delta \vdash \chi \wedge \neg\chi$. $(\Diamond 1)$ yields $\Diamond(\chi \wedge \neg\chi) \in \Gamma$, and thus by R19: $\chi \wedge \neg\chi \in \Gamma$, contradicting consistency of Γ.
3. Δ is saturated, for assume that $\chi \vee \chi' \in \Delta$, then, by the special construction of the sequence $\{\varphi_n\}_n$, $\chi \vee \chi' = \varphi_n$ and $\Delta_n \vdash \varphi_n$ for some n. By definition, Δ_{n+1} contains χ or χ', and so does Δ.[8]

[8] One might object at this point that the definition seems to prohibit containment of *both* disjuncts. However, this objection fails for at least two reasons: (i) it may, for example, be the case that $\chi' \in \Delta_n$ while $\Delta_n \cup \{\chi\}$ has $(\Diamond 1)$, and then $\{\chi, \chi'\} \subseteq \Delta_{n+1}$; (ii) in case $\chi \notin \Delta_n$ and $\chi' \notin \Delta_n$, whereas both $\Delta_n \cup \{\chi\}$ and $\Delta_n \cup \{\chi'\}$ have $(\Diamond 1)$, the definition yields that only $\chi \in \Delta_{n+1}$ but $\chi' \notin \Delta_{n+1}$. However, the equivalent $\chi' \vee \chi$ will eventually turn up as φ_k, and then possibly $\chi' \in \Delta_{k+1}$.

4. $\varphi \in \Delta_0 \subseteq \Delta$.
5. $\Gamma\mathcal{R}\Delta$ since
   - Δ is a CST,
   - $\Box\delta \in \Gamma$ implies $\delta \in \Delta_0 \subseteq \Delta$,
   - $\delta \in \Delta$ implies $\Diamond\delta \in \Gamma$ by ($\Diamond$1). ■

Lemma 4.3 *$\Box\varphi \in \Gamma$ iff for all Δ such that $\Gamma\mathcal{R}\Delta$: $\varphi \in \Delta$.*

Proof: From the left to the right this follows immediately from the definition of $\mathcal{R}$. To show the other direction, we argue by contraposition. Suppose that $\Box\varphi \notin \Gamma$. We have to construct a CST Δ such that $\Gamma\mathcal{R}\Delta$ and $\varphi \notin \Delta$. Let $\varphi_0, \varphi_1, \dots, \varphi_n \dots$ be an enumeration with countable repetition of the formulas of the standard modal language. Δ_n is defined in such a way that it respects the *second diamond-property* ($\Diamond$2). Σ has $\Diamond$2 if for all ε:

$$(\Diamond 2) \qquad \Sigma \vdash \varepsilon \vee \varphi \Rightarrow \Diamond\varepsilon \in \Gamma.$$

Δ_n is defined recursively by:

- $\Delta_0 = \Box^{-1}[\Gamma] = \{\delta \mid \Box\delta \in \Gamma\}$;
- $\Delta_{n+1} = \Delta_n$ if $\Delta_n \not\vdash \varphi_n$;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n\}$ if $\Delta_n \vdash \varphi_n$ and φ_n is not a disjunction;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ if $\Delta_n \vdash \varphi_n$, $\varphi_n = \chi \vee \chi'$ and $\Delta_n \cup \{\chi\}$ has ($\Diamond$2);
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi'\}$ if $\Delta_n \vdash \varphi_n$, $\varphi_n = \chi \vee \chi'$ and $\Delta_n \cup \{\chi\}$ does not have ($\Diamond$2).

Let Δ be $\bigcup_n \Delta_n$. First we will check that the Δ_n and Δ share ($\Diamond$2); next that Δ is a CST with the desired properties. That Δ_n has the $\Diamond$2-property is shown by induction on n:

- Let $\Delta_0 \vdash \varepsilon \vee \varphi$, then (R10) there are $\delta_1, \dots, \delta_m \in \Delta_0$ such that $\delta \vdash \varepsilon \vee \varphi$, where $\delta = \delta_1 \wedge \dots \wedge \delta_m$. By R15 and R18 this implies $\Box\delta \vdash \Diamond\varepsilon \vee \Box\varphi$ (2). Since $\Box\delta_i \in \Gamma$, R13 and closure of CSTs under conjunction, we obtain $\Box\delta \in \Gamma$, consequently with (2): $\Diamond\varepsilon \vee \Box\varphi \in \Gamma$. So, by saturation and the fact that $\Box\varphi \notin \Gamma$: $\Diamond\varepsilon \in \Gamma$. Whence Δ_0 has ($\Diamond$2).
- Suppose Δ_n has $\Diamond$2 (IH), and consider Δ_{n+1}.
  - In case $\Delta_n \not\vdash \varphi_n$ there is nothing to prove.
  - If $\Delta_n \vdash \varphi_n$, φ_n not a disjunction, let $\Delta_{n+1} \vdash \varepsilon \vee \varphi$, i.e. $\Delta_n, \varphi_n \vdash \varepsilon \vee \varphi$. The cut-rule then provides $\Delta_n \vdash \varepsilon \vee \varphi$ and so by IH: $\Diamond\varepsilon \in \Gamma$.
  - If $\Delta_n \vdash \varphi_n = \chi \vee \chi'$ and $\Delta_n \cup \{\chi\}$ has ($\Diamond$2), so has $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ (apply cut).
  - Let $\Delta_n \vdash \varphi_n$, $\varphi_n = \chi \vee \chi'$ and for some ε_1: $\Delta_n, \chi \vdash \varepsilon_1 \vee \varphi$ and $\Diamond\varepsilon_1 \notin \Gamma$. Then $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi'\}$ has ($\Diamond$2). For suppose on the contrary that there exists an ε_2 such that $\Delta_n, \varphi_n, \chi' \vdash \varepsilon_2 \vee \varphi$ while $\Diamond\varepsilon_2 \notin \Gamma$. By R5 and R6 (including associativity of $\vee$), $\Delta_n, \varphi_n, \chi \vee \chi' \vdash \varepsilon_1 \vee \varepsilon_2 \vee \varphi$, thus by cut: $\Delta_n \vdash \varepsilon_1 \vee \varepsilon_2 \vee \varphi$, so (IH) $\Diamond(\varepsilon_1 \vee \varepsilon_2) \in \Gamma$. With R14, $\Diamond\varepsilon_1 \vee \Diamond\varepsilon_2 \in \Gamma$, and by saturation of Γ: $\Diamond\varepsilon_1 \in \Gamma$ or $\Diamond\varepsilon_2 \in \Gamma$, contradictory to the assumptions.

So each Δ_n has the property ($\Diamond 2$). Consequently, Δ also has ($\Diamond 2$). The rest of the proof is straightforward.

1. Δ is a theory: the proof is the same as in the previous lemma.
2. $\varphi \notin \Delta$, for assume on the contrary that $\varphi \in \Delta$. So $\Delta \vdash (\varphi \wedge \neg\varphi) \vee \varphi$, and by ($\Diamond 2$): $\Diamond(\varphi \wedge \neg\varphi) \in \Gamma$, and therefore (R19) $\Box\varphi \in \Gamma$, which contradicts $\Box\varphi \notin \Gamma$.
3. This also shows that Δ is consistent (by R8).
4. Δ is saturated. The argument is completely similar to the one spelled out in the previous lemma.
5. $\Gamma\mathcal{R}\Delta$ since
   - Δ is a CST,
   - $\Box\delta \in \Gamma$ implies $\delta \in \Delta_0 \subseteq \Delta$,
   - $\delta \in \Delta \Rightarrow \Delta \vdash \delta \vee \varphi \Rightarrow$ ($\Diamond 2$) $\Diamond\delta \in \Gamma$

■

We are ready to prove the truth/falsity lemma for the canonical model defined in the proof of theorem 4.2.

Lemma 4.4 *For all* $\Gamma \in \mathcal{S}$: $\Gamma \models \psi$ *iff* $\psi \in \Gamma$; $\Gamma \dashv\!\!\mid \psi$ *iff* $\neg\psi \in \Gamma$.

Proof: by induction on the structure of ψ.

- if ψ is a propositional atom, the lemma holds by definition of $\mathcal{V}$.

  Assume the lemma to hold for ψ and χ (IH).
- $\Gamma \models \neg\psi$ iff $\Gamma \dashv\!\!\mid \psi$ iff (IH) $\neg\psi \in \Gamma$.
  $\Gamma \dashv\!\!\mid \neg\psi$ iff $\Gamma \models \psi$ iff (IH) $\psi \in \Gamma$ iff (R1) $\neg\neg\psi \in \Gamma$.
- $\Gamma \models \psi \wedge \chi$ iff $\Gamma \models \psi$ & $\Gamma \models \chi$ iff (IH) $\psi \in \Gamma$ & $\chi \in \Gamma$ iff (R4,R10) $\psi \wedge \chi \in \Gamma$.
  $\Gamma \dashv\!\!\mid \psi \wedge \chi$ iff $\Gamma \dashv\!\!\mid \psi$ or $\Gamma \dashv\!\!\mid \chi$ iff (IH) $\neg\psi \in \Gamma$ or $\neg\chi \in \Gamma$ iff (R5 and saturation) $\neg\psi \vee \neg\chi \in \Gamma$ iff (R2) $\neg(\psi \wedge \chi) \in \Gamma$.
- $\Gamma \models \Box\psi$ iff for all Δ such that $\Gamma\mathcal{R}\Delta$: $\Delta \models \psi$ iff (IH) for all Δ such that $\Gamma\mathcal{R}\Delta$: $\psi \in \Delta$ iff (lemma 4.3) $\Box\psi \in \Gamma$.
  $\Gamma \dashv\!\!\mid \Box\psi$ iff for some Δ such that $\Gamma\mathcal{R}\Delta$: $\Delta \dashv\!\!\mid \psi$ iff (IH) for some Δ such that $\Gamma\mathcal{R}\Delta$: $\neg\psi \in \Delta$ iff (lemma 4.2) $\Diamond\neg\psi \in \Gamma$ iff (R11) $\neg\Box\psi \in \Gamma$.

The steps for $\vee$ and $\Diamond$ are skipped, since we can redefine these in terms of $\neg$, $\wedge$ and $\Box$. This reduction was noticed before to be legitimate on the semantic side; now on the syntactic side, it is licensed by R1, R2, R11, R15, etcetera. ■

4.3.2 Modal falsifiability

Since for propositional formulas falsifiable validity on coherent situations leads to classical **pL**, one expects an analogous result for modal formulas. So, is **K** determined by FALSIF on coherent models? And, similarly for abolute rules: is $\mathbf{K_a}$ sound and complete for absolute falsifiable validity on coherent models?

A prerequisite for generalizing the proof of theorem 3.3 is a reinspection of some properties discussed in sections 3.2 and 4.3. We noticed that in general *persistence* does not hold for modal logic. We seem to be in big trouble here: the proof of theorem 3.3 depended on persistence. But on second thoughts there is no reason to panic: model *completion* satisfies external persistence which also holds for the modal case.

Theorem 4.3 *For modal logic the coherent semantics with absolute falsifiable validity is complete with respect to modal system* $\mathbf{K_a}$.

Proof: similar to the proof of theorem 3.3, by using *inherited classicality* and *external persistence* (see proposition 4.3). ∎

For *mixed* falsifiable validity, we notice that van Benthem's reinterpretation of Beth tableaus easily generalizes to modal logic.

Theorem 4.4 FALSIF$_{mix}$ *on partial Kripke models is completely described by the modal logic* $\mathbf{K_r}$.

Proof: since FALSIF$_{mix}$ validity of φ/ψ is equivalent to $\not\dashv \varphi \rightarrow \psi$, which is described by $\mathbf{K}$, proposition 4.1 provides the suitable relational format. ∎

Since the frame is not effected by the proof procedure, the last two theorems allow an obvious extension to normal systems:

Theorem 4.5
For falsifiable validity on coherent models the (frame complete) normal modal systems are captured by the usual conditions on accessibility.

So the tenor of all this is similar to that for the purely propositional case: for formulas there is no descriptive difference between standard Kripke semantics and partial semantics under the 'never false' concept of validity. Similarly, for rules classical and partial Kripke semantics (under the *mixed* version of relative falsifiable validity) also amounts to the same. Yet, there may still be important differences between both approaches. One may be that of the greater intuitive ('realistic') appeal of situations. Another the greater computational efficiency of the partial approach: the states of the partial models can now be specified by considerably smaller sets. Moreover, a partial model nonfalsifying a formula may contain less states than the classical model doing the same job.

Apart from a relativized formulation, we need some more modifications to describe (straight) relative non-falsification. Similar to considerations leading to theorem 3.4 in chapter 3 we find that the *ex falso* principles are not validated anymore. So, compared to $\mathbf{M}^+$, R8 and R19 are out. Let $\mathbf{M}^*$ be the system where R8 and R19 are replaced by R8* and R19* (which may be thought as to result from the *ex falso* rules by contraposition):

(R8*) $\psi \vdash \varphi \vee \neg\varphi$ (*tertium non datur*)

(R19*) $\psi \vdash \Box(\varphi \vee \neg\varphi)$ *(modal tertium non datur)*

Theorem 4.6 *On coherent models* FALSIF$_{rel}$ *is sound and complete with respect to the modal system* **M***.

The proof of this is postponed to the next section.

4.4 Situational modal models

Releasing the restriction to coherence, the Kripke frame can be supplemented with a partial, possibly incoherent valuation. We arrive at the notion of a modal situation model.

Definition 4.1 (modal situation model) *A structure* $\langle S, R, V\rangle$ *is a modal situation model if* $R \subseteq S \times S$ *and* $V : Prop \times S \longrightarrow \wp(\{0,1\})$.

The truth conditions are as before, i.e. the general situation clauses for atoms, the standard ones for the connectives and the partial Kripkean ones for modal operators. Moreover, the definitions for the options with respect to validity and rules are unchanged.

There is, however, a clear difference on the issue of locally defined conditions such as *coherence* and *persistence*. As we saw in section 4.3 the global versions of these notions (for coherent *models* and *external* extensions) generalize to complex formulas. As we already saw for persistence, the local variants of these conditions (i.e. coherent *situations*, etcetera) do not allow for such a generalization, due to the essentially global nature of Kripke models. So, a coherent situation can be incoherent with respect to a modal formula if there exists an accessible incoherent situation. The same holds, *mutatis mutandis* for *totality* and *duality*. What we can do, of course, is to redefine the notions of *total model* and *duality* for situational Kripke semantics. We omit these redefinitions since they are obviously obtained from those in section 3.3, by decorating the models with an accessibility relation.

Presumably because of the global nature of validity, the propositional and modal case are quite similar for general situation models. To begin with, the number of possible systems allowed by the different types of validity and rules is immediately reduced by the following two facts:

- there are still no valid formulas in this semantics (cf. theorem 3.5);
- duality preservation also holds for the standard modal language (cf. proposition 3.14).[9]

Consequently, the useful reduction of falsifiability to verification expressed earlier in proposition 3.15 is maintained. Moreover, by their very nature, *absolute* and *mixed*

[9]See [Th90a, p.48,63]

verification of rules produce the trivial rule systems $\mathcal{L} \times \mathcal{L}$ (for inferences with one premise) and $\emptyset$, respectively. So, what remains to be inspected is *relative verification.*

Compared to coherent models we notice that neither the pair R8, R19 nor the pair R8*, R19* hold. Analogous to the propositional case elimination of these rules supplies a complete set of rules. Thus let **M**, the modal counterpart of relevance logic, be generated by the description of $\mathbf{M}^+$ minus rules R8 and R19.

Theorem 4.7
The modal logic for relative validity on situation models is **M**.

Proof: The completeness proof is a modification of that for theorem 4.2. Since the canonical situations now may be syntactically inconsistent sets, we consider *saturated theories* (STs) instead of CSTs. The canonical valuation function also needs some modification, since it can be multi-valued. So, the canonical model $\mathcal{M} = \langle \mathcal{S}, \mathcal{R}, \mathcal{V} \rangle$ is defined by:

- $\mathcal{S} = \{\Gamma \mid \Gamma \text{ is an ST}\}$;
- $\Gamma \mathcal{R} \Delta$ iff
  - $\Gamma, \Delta \in \mathcal{S}$,
  - $\Box\psi \in \Gamma$ implies $\psi \in \Delta$ for all ψ, and
  - $\psi \in \Delta$ implies $\Diamond\psi \in \Gamma$ for all ψ;
- $1 \in \mathcal{V}(p, \Gamma)$ iff $p \in \Gamma$, and $0 \in \mathcal{V}(p, \Gamma)$ iff $\neg p \in \Gamma$.

Suppose $\Sigma \nvdash \varphi$. Notice that the Lindenbaum lemma, the truth/falsity lemma and lemma 4.2 hold for STs: R8 and R19 were only used for proving consistency. The counterpart of lemma 4.3 is more complicated, since the subproof that $\varphi \notin \Delta$ uses R19. This lemma is reproven for STs and **M** below. Then by the Lindenbaum lemma (with respect to **M**) Σ can be extended to an ST Ω such that $\varphi \notin \Omega$. Since $\Sigma \subseteq \Omega$ and $\varphi \notin \Omega$, the truth/falsity lemma gives $\mathcal{M}, \Omega \models \Sigma$ and $\mathcal{M}, \Omega \not\models \varphi$, whence $\Sigma \not\models \varphi$. ∎

The remedy for the complication concerning the counterpart of lemma 4.3 is to build the desired fact ($\varphi \notin \Delta$) into the definition of Δ. This can be done without using either R8 or R19.

Lemma 4.5 $\Box\varphi \in \Gamma$ *iff for all* ST Δ *such that* $\Gamma\mathcal{R}\Delta$*:* $\varphi \in \Delta$.

Proof: From the left to the right this is trivial. The other direction is shown by an indirect proof. Assume that $\Box\varphi \notin \Gamma$.

Let $\{\varphi_n\}_n$ be an enumerating sequence of formulas with countable repetition. Δ_n is defined in such a way that it respects $\Diamond 2$ without containing φ:

- $\Delta_0 = \{\delta \mid \Box\delta \in \Gamma\}$;
- $\Delta_{n+1} = \Delta_n$ if $\Delta_n \nvdash \varphi_n$;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n\}$ if $\Delta_n \vdash \varphi_n$ and φ_n is not a disjunction;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \psi\}$ if $\Delta_n \vdash \varphi_n$, $\varphi_n = \psi \vee \chi$, $\Delta_n \cup \{\psi\}$ has $\Diamond 2$ and $\Delta_n \cup \{\psi\} \nvdash \varphi$.

- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ if $\Delta_n \vdash \varphi_n$, $\varphi_n = \psi \vee \chi$, and either $\Delta_n \cup \{\psi\}$ does not have $\Diamond 2$ or $\Delta_n \cup \{\psi\} \vdash \varphi$.

Again $\Delta = \bigcup_n \Delta_n$. The proof of theoremhood of Δ is still identical; saturation is quite trivial again. This leaves two steps to be spelled out:

1. That Δ_n and thus Δ obey $\Diamond 2$ is shown by an inductive proof similar to that for lemma 4.3; in fact the only case that needs reconsideration is that in which $\Delta_n \vdash \varphi_n$ and φ_n is a disjunction, say $\varphi_n = \psi \vee \chi$. There are but two possibilities:
   - $\Delta_n \cup \{\psi\}$ has $\Diamond 2$, and $\Delta_n \cup \{\psi\} \not\vdash \varphi$. Thus, by application of the cut rule $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \psi\}$ has $\Diamond 2$;
   - either $\Delta_n \cup \{\psi\}$ does not have $\Diamond 2$, or $\Delta_n \cup \{\psi\} \vdash \varphi$ (or both).
     - If $\Delta_n \cup \{\psi\}$ does not have $\Diamond 2$ then by a reasoning analogous to lemma 4.3, $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ has $\Diamond 2$.
     - If $\Delta_n \cup \{\psi\} \vdash \varphi$, suppose that $\Delta_n \cup \{\varphi_n, \chi\}$ does not have $\Diamond 2$. So there is an ε such that $\Delta_n, \varphi_n, \chi \vdash \varepsilon \vee \varphi$, and $\Diamond\varepsilon \notin \Gamma$. However, by R5 and R10 we also have $\Delta_n, \varphi_n, \psi \vdash \varepsilon \vee \varphi$, and thus (R6): $\Delta_n, \varphi_n, \vdash \varepsilon \vee \varphi$, and by cut: $\Delta_n \vdash \varepsilon \vee \varphi$, and IH yields $\Diamond\varepsilon \in \Gamma$. Contradiction. Thus Δ_{n+1} has $\Diamond 2$.

2. $\varphi \notin \Delta$ follows since it can be shown by induction that $\Delta_n \not\vdash \varphi$:
   - For $n = 0$, suppose on the contrary that $\Delta_0 \vdash \varphi$. By R10 there are $\delta_1, \ldots, \delta_m \in \Delta_0$ such that $\delta_1 \wedge \cdots \wedge \delta_m \vdash \varphi$, so (R15) $\Box\delta \vdash \Box\varphi$ where $\delta = \delta_1 \wedge \cdots \wedge \delta_m$. Since $\Box\delta \in \Gamma$ it follows that $\Box\varphi \in \Gamma$. Contradiction.
   - Suppose $\Delta_n \not\vdash \varphi$ (IH). Consider Δ_{n+1}.
     - In case $\Delta_n \not\vdash \varphi_n$: $\Delta_{n+1} = \Delta_n \not\vdash \varphi$.
     - If $\Delta_n \vdash \varphi_n$, and φ_n is not a disjunction, suppose $\Delta_{n+1} \vdash \varphi$, i.e. $\Delta_n, \varphi_n \vdash \varphi$. Then by the cut rule: $\Delta_n \vdash \varphi$, which contradicts IH.
     - if $\Delta_n \vdash \varphi_n$, $\varphi_n = \psi \vee \chi$, $\Delta_n \cup \{\psi\}$ has $\Diamond 2$ and $\Delta_n \cup \{\psi\} \not\vdash \varphi$, then (cut) $\Delta_{n+1} \not\vdash \varphi$.
     - If $\Delta_n \cup \{\psi\}$ does not have $\Diamond 2$, assume that $\Delta_{n+1} \vdash \varphi$, i.e. $\Delta_n, \varphi_n, \chi \vdash \varphi$. There is an ε such that $\Delta_n, \varphi_n, \psi \vdash \varepsilon \vee \varphi$, and $\Diamond\varepsilon \notin \Gamma$. However, by R5 also $\Delta_n, \varphi_n, \chi \vdash \varepsilon \vee \varphi$, and thus (R6): $\Delta_n, \varphi_n, \vdash \varepsilon \vee \varphi$, and by cut: $\Delta_n \vdash \varepsilon \vee \varphi$, and IH yields $\Diamond\varepsilon \in \Gamma$. Contradiction.
     - If $\Delta_n, \psi \vdash \varphi$, $\Delta_n, \varphi_n, \chi \not\vdash \varphi$ by cut and R5.

Finally the step establishing $\Gamma \mathcal{R} \Delta$ is as before. ■

With this result another route to our earlier proof for $\mathbf{M}^+$ becomes available: show for lemmas 4.2 and 4.3 that the addition of axioms R8 and R19 brings about the consistency of Δ, given the consistency of Γ. We avoided such a reformulation for expository reasons: the earlier proof is surely more transparent.

By the same strategy as in the propositional case, *duality* may be used to prove theorem 4.6. Duality boils down to interchanging true and not-false, and false and not-true. Here its effect is a reduction of relative falsification on coherent models to relative verification on total ones.

Theorem 4.6 (repeated) *On coherent models* FALSIF$_{rel}$ *is sound and complete with respect to the modal system* **M***.

Proof: As indicated this reduces to showing that **M*** is sound and complete for VERIF$_{rel}$ on total models. Soundness is straightforward, and completeness amounts to showing that the canonical model constructed for theorem 4.7 is total: the canonical situations are *full* (i.e. FSTs), which for **M*** boils down to non-empty STs.

First notice that the main proof and the truth lemma itself are not affected by fullness. Next, for the Lindenbaum lemma the construction is independent of the extra rules; since $\Sigma \neq \emptyset$, R8* immediately shows fullness. This leaves the lemmas 4.2 and 4.5 to be reconsidered. For lemma 4.2 note that $\varphi \in \Delta$, so $\psi \vee \neg\psi \in \Delta$ by R8. For lemma 4.5, we note that Γ is an FST, therefore $\Gamma \neq \emptyset$, so by R19* $\Box(\psi \vee \neg\psi) \in \Gamma$, thus by the construction of Δ: $\psi \vee \neg\psi \in \Delta$. Hence Δ is also an FST. ■

4.5 Alternatives

In the literature one can find other proposals, suggesting to vary the truth conditions, the notion of validity and the accessibility relation. Some of these proposals are rather eclectic or even *ad hoc* in nature. Here we will consider only two alternatives to our standard theory.[10]

4.5.1 extending the language

Reconsider the extended languages from chapters 1 and 2, and the remarks on validity and completeness in section 3.4.1. It follows from the modal extension of the Langholm reduction of partial to classical logic [11] that the $\neg$–free fragment of the extended language is essentially classical as far as validity is concerned:

Proposition 4.4 *If* $\varphi \in \mathcal{L}_{\sim,\wedge,\vee,\Box,\Diamond}$ *then* $\models\!\!\!\equiv \varphi \Leftrightarrow\, \models \varphi$.

Proof: For the $^+/^-$ translation, see section 2.4. Notice that $^+$ can now be defined without reference to $^-$. Replacing output $\neg$ by $\sim$ implies that $^+$ does not effect the formula. Then by proposition 2.4 it follows that trivalent or quadrivalent truth can be transformed to bivalent truth, and *vice versa*. ■

For the extended modal language with ∂ instead of $\sim$ we obtain a similar result.[12]

Proposition 4.5 *If* $\varphi \in \mathcal{L}_{\partial,\wedge,\vee,\Box,\Diamond}$ *then* $\models\!\!\!\equiv \varphi \Leftrightarrow\, \models \varphi$.

For the unrestricted extensions (i.e. the standard languages *including* $\neg$, conjoined with either $\sim$ or ∂) there are similar though more complicated reductions of partial

[10] [Th90b] also considers Levesque's logic of explicit belief from [Le84a], which differs in many respects from our standard logic. See also chapter 7 of this thesis.

[11] *Vide* chapter 2.

[12] The Langholm $^+/^-$ translation clauses for ∂ are $(\partial\varphi)^+ = \neg\varphi^+$ and $(\partial\varphi)^- = \neg\varphi^-$.

to classical logic. An extended formula over the propositional variables $p_1, \ldots, p_n$ is translated into a standard formula over $p_1^+, p_1^-, \ldots, p_n^+, p_n^-$. The extended formula is partially valid iff the standard formula resulting from the translation is classically valid. Since we will restrict ourselves mostly to the standard language in the rest of this thesis, we will not go any further on this here.[13]

4.5.2 partial accessibility

In [Mu89] various modal operators connected to verbs such as 'know' and 'believe' are analysed in an intensional type logic. The analysis may be conceived as to formalize a modeltheoretic interpretation. It is obtained by a compositional translation that partializes virtually every relation. Consequently, accessibility is given the same (partial) treatment as other predicates. This results in splitting the accessibility relation.[14]

Now $s \models \Box\varphi$ amounts to $[\![\lambda i \forall j(Rij \rightarrow \varphi(j))]\!](s) = 1$ and $s =\!\!| \Box\varphi$ likewise to $[\![\lambda i \forall j(Rij \rightarrow \varphi(j))]\!](s) = 0$. Then by the (partial) interpretation rules for application, λ-abstraction, quantification, predication, and the clauses for $\rightarrow$ stated in section 3.2, this leads to the following truth/falsity conditions for $\Box$:

$$\begin{array}{ll} s \models \Box\varphi \Leftrightarrow & \text{for every } t : s[\![R]\!]^- t \text{ or } t \models \varphi \\ s =\!\!| \Box\varphi \Leftrightarrow & \text{for some } t : s[\![R]\!]^+ t : \text{ and } t =\!\!| \varphi \end{array}$$

To simplify the formulation, let $([\![R]\!]^-)^c = \mathsf{R}$ and $[\![R]\!]^+ = \mathsf{R}'$ [15], then

$$\begin{array}{ll} s \models \Box\varphi \Leftrightarrow \forall t \in \mathsf{R}[s] : t \models \varphi & s =\!\!| \Box\varphi \Leftrightarrow \exists t \in \mathsf{R}'[s] : t =\!\!| \varphi \\ s \models \Diamond\varphi \Leftrightarrow \exists t \in \mathsf{R}'[s] : t \models \varphi & s =\!\!| \Diamond\varphi \Leftrightarrow \forall t \in \mathsf{R}[s] : t =\!\!| \varphi \end{array}$$

So a model M is now of the form $\langle S, \mathsf{R}, \mathsf{R}', V\rangle$. For *coherent* models we need a restriction on admissible frames: only models with $\mathsf{R}' \subseteq \mathsf{R}$ are allowed. With respect to these models, propositional coherence entails modal coherence.

We can also give strong completeness theorems that are intimately connected to the earlier characterizations for the standard cases. Consider the deductive system $\mathbf{M}^{+-} = \mathbf{M}^{+} - \{\text{R18}\}$.

Theorem 4.8 *With partial accessibility relations,* VERIF$_{rel}$ *on coherent models is sound and complete for the system* $\mathbf{M}^{+-}$.

Proof: analogous to theorem 4.2, but somewhat easier. Let S be the set of all CSTs with respect to $\mathbf{M}^{+-}$. The definition of $\mathcal{R}'$ is still 'two-fold', but that of $\mathcal{R}$ is 'single':

- $\Gamma\mathcal{R}\Delta$ iff $\Box\varphi \in \Gamma$ implies $\varphi \in \Delta$ for all φ;
- $\Gamma\mathcal{R}'\Delta$ iff $\Box\varphi \in \Gamma$ implies $\varphi \in \Delta$, and $\varphi \in \Delta$ implies $\Diamond\varphi \in \Gamma$ for all φ.

[13] Cf. [FH*87] for validity on an extended modal language, with $\Box$ interpreted by means of $\sqsubseteq$.

[14] Within Levesque's approach (cf. footnote 10), [La87] also has split accessibility, but different (non-standard) truth conditions and validity.

[15] $[\![R]\!]^+$ is the so-called *(positive) extension* of R (=set of verifying situations), and $[\![R]\!]^-$ the *negative* (or, *anti-*)*extension* of R (=set of falsifying situations).

The truth lemma then holds since lemma 4.2 refers to the same two-fold canonical relation and does not use R18; the counterpart of lemma 4.3 is easily proved, since the $\Diamond 2$ property is not needed anymore. To be more specific, define Δ_0 and Δ_{n+1} as before, apart from the case of disjunction which is:

- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \psi\}$ if $\varphi_n = \psi \vee \chi$, $\Delta_n \vdash \varphi_n$, and $\Delta_n, \psi \nvdash \varphi$;
- $\Delta_{n+1} = \Delta_n \cup \{\varphi_n, \chi\}$ if $\varphi_n = \psi \vee \chi$, $\Delta_n \vdash \varphi_n$, and $\Delta_n, \chi \nvdash \varphi$.

Theoremhood and non-containment of φ are straightforward. Saturation is likewise unproblematic. Observe that a lot of properties directly follow from the fact that $\mathbf{M}^{+-}$ contains all of $\mathbf{rL}^+$ and from the fact that the construction of Δ resembles the one in the proof theorem 3.1. ∎

Once we allow incoherent situations, as advocated by Muskens, the $\subseteq$-order between the accessibility relations can be entirely dismissed. The proof system is consequently smaller: let $\mathbf{M}^{--}$ be $\mathbf{M} - \{$ R17,R18$\}$.

Theorem 4.9 *With partial accessibility relations,* VERIF$_{rel}$ *on situational models is sound and complete for the system* $\mathbf{M}^{--}$.

Proof: still analogous, and still easier. For the canonical model let S be the set of all STs with respect to $\mathbf{M}^{--}$. The canonical $\mathcal{R}$ and $\mathcal{R}'$ are now both defined by 'single' clauses:

- $\Gamma \mathcal{R} \Delta$ iff $\Box\varphi \in \Gamma$ implies $\varphi \in \Delta$ for all φ;
- $\Gamma \mathcal{R}' \Delta$ iff $\varphi \in \Delta$ implies $\Diamond\varphi \in \Gamma$ for all φ.

The separated clauses avoid interference of $\Box$ with $\Diamond$ in the sublemmas, which trivializes the proof. ∎

Although the general version of Muskens's system is elegant and indicates its feasibility of being part of a typed partial logic, we feel that the coherent version, which should be preferable for modelling epistemic attitudes, seems technically less elegant (cf. the asymmetrical nature of the canonical accessibility relation). More importantly, we do not have intuitions that support the proposed splitting of the accessibility relation.

4.6 Special systems

We will now focus on some systems of particular interest for applications on knowledge and belief. The usual attributes of knowledge and belief concern *veridicality* (truth of knowledge, consistency of belief) and *introspection*. So our quest is for the partial counterparts of normal modal systems such as **T**, **NKD4**, **S4** and **S5**.[16] For the sake of computational use of knowledge, we are especially interested in some properties of these systems that guarantee finite representation, in particular the finite model property, decidability and logical finiteness. To achieve these we need a few additional techniques: filtration and normal forms.

[16]The *system* **T** is characterized by the axioms and rules **pL**, **N**, **K** and **T**, i.e. as **NKT** (extending the so-called Lemmon code with 'bf N'). Likewise, we have **S4**=**NKT4** and **S5**=**NKT5**.

4.6.1 Completeness results

Standard normal systems of epistemic and doxastic logic share some or all of the following properties:

(D) $\vdash \Box\varphi \to \Diamond\varphi$;

(T) $\vdash \Box\varphi \to \varphi$;

(4) $\vdash \Box\varphi \to \Box\Box\varphi$;

(5) $\vdash \Diamond\varphi \to \Box\Diamond\varphi$;

In a cognitive interpretation these axioms are claimed to express consistency of knowledge and belief (**D**), truth of knowledge (**T**), and positive (**4**) and negative (**5**) introspection (self-reflection) of knowledge and belief. In many applications these principles are indeed defensible. However, the normal systems they lead to (for example, **T**, **S4**, **S5**) suffer from problems of logical omniscience (cf. chapter 6). In our view, these problems do not arise because of the axioms above, but because of the core logic **K**. As we saw earlier, the system **K** is not not valid on all partial models.

So the question arises whether the listed properties of cognitive operators can also be captured in partial logic without reintroducing (all types of) logical omniscience again. Since we do not have tautologies according to the preferential types of validity and models (viz. verification on coherent or arbitrary situations), the determining principles take the form of primitive deduction rules rather than axioms. These rules can then be added to the 'core systems' **M**, $\mathbf{M}^+$ and $\mathbf{M}^*$, which correspond to the different notions of consequence (verification on arbitrary, coherent or total situations, respectively). The proposed rules are (with obvious nomenclature):

($\mathbf{D_r}$) $\Box\varphi \vdash \Diamond\varphi$;

($\mathbf{T_r}$) $\Box\varphi \vdash \varphi \;\;\&\;\; \varphi \vdash \Diamond\varphi$;

($\mathbf{4_r}$) $\Box\varphi \vdash \Box\Box\varphi \;\;\&\;\; \Diamond\Diamond\varphi \vdash \Diamond\varphi$;

($\mathbf{5_r}$) $\Diamond\varphi \vdash \Box\Diamond\varphi \;\;\&\;\; \Diamond\Box\varphi \vdash \Box\varphi$;

Notice that apart from $\mathbf{D_r}$ the rules come in dual pairs. The reason for this is that for $\mathbf{M}^+$ and $\mathbf{M}^*$ the meta-rule of contraposition is not valid in general, but appears to hold for the listed principles and therefore has to be stipulated. $\mathbf{D_r}$ is simply self-dual.

In normal modal logic **D**, **T**, **4** and **5** correspond to the following structural conditions on frames:

seriality: for all $s \in S$ there is a $t \in S$ such that sRt;

reflexivity: for all $s \in S$: sRs;

transitivity: for all $s, t, u \in S$: $sRt, tRu \Rightarrow sRu$;

euclidicity: for all $s, t, u \in S$: $sRt, sRu \Rightarrow tRu$.

We are now in the position to state and prove the completeness of the partial counterparts of standard normal systems.

Theorem 4.10 (strong completeness for special systems)
Given the corresponding classes of models for the systems $\mathbf{M}, \mathbf{M}^+$, *and* $\mathbf{M}^*$ *(containing arbitrary, coherent and total situations, respectively), addition of the following rules is captured by structural constraints:*

$\mathbf{D_r}$ *is sound and complete for serial models*

$\mathbf{T_r}$ *is sound and complete for reflexive models*

$\mathbf{4_r}$ *is sound and complete for transitive models*

$\mathbf{5_r}$ *is sound and complete for euclidean models*

Proof: soundness is straightforward and is mostly left to the reader. For example, consider the case of $\mathbf{5_r}$. Suppose that M is a euclidean model such that for arbitrary φ, s: $M, s \models \Diamond\varphi$. So there is a t such that sRt (1) and $M, t \models \varphi$ (2). Then for all u with sRu (by euclidicity and (1)) uRt, and so (by (2)) $M, u \models \Diamond\varphi$, thus $M, s \models \Box\Diamond\varphi$. Therefore $\Diamond\varphi \models \Box\Diamond\varphi$. The contrapositive is similar.

Completeness is proved by a Henkin-type proof, cf. sections 4.3 and 4.4. It suffices to show that the canonical model $\mathcal{M}$ for system $\mathbf{S}$ has the structural property. Recall that the canonical situations are *saturated theories* for $\mathbf{M}$, *consistent saturated theories* for $\mathbf{M}^+$, and *full saturated theories* for $\mathbf{M}^*$. Moreover, canonical accessibility $\mathcal{R}$ is triggered by the elegant

$$\Gamma\mathcal{R}\Delta \Leftrightarrow \Box^{-1}\Gamma \subseteq \Delta \subseteq \Diamond^{-1}\Gamma.$$

The canonical interpretation is given by

$$\mathcal{V}(p, \Sigma) \ni \begin{cases} 1 & \text{if} \quad p \in \Sigma \\ 0 & \text{if} \quad \neg p \in \Sigma \end{cases}$$

We can now treat the separate items, the first one being a bit harder than the others.

($\mathbf{D_r}$) Suppose that Γ is an ST (CST, FST) with respect to $\mathbf{MD_r}$ ($\mathbf{M^+D_r}$, $\mathbf{M^*D_r}$). Distinguish between the following cases:

1. $\Diamond\varphi \in \Gamma$ for some φ. So by lemma 4.2, there is an ST (CST, FST) Δ such that $\Gamma\mathcal{R}\Delta$ and $\varphi \in \Delta$.
2. $\Diamond\varphi \notin \Gamma$ for all φ. Since $\mathbf{S}$ is an $\mathbf{MD_r}$–theory it cannot contain any formula $\Box\varphi$ either, so $\Box^{-1}\Gamma = \Diamond^{-1}\Gamma = \emptyset$. Now if $\mathbf{S}$ does not contain $\mathbf{M}^*$, we may choose $\Delta = \emptyset$, which is a CST. If $\mathbf{S}$ contains $\mathbf{M}^*$, this case cannot occur. For $\Gamma \neq \emptyset \Rightarrow \gamma \in \Gamma$ for some γ, and so since Γ is an $\mathbf{M}^*$–theory: $\Box(\gamma \vee \neg\gamma) \in \Gamma$, thus ($\mathbf{D_r}$) $\Diamond(\gamma \vee \neg\gamma) \in \Gamma$.

Therefore $\mathcal{M}$ is serial.

($\mathbf{T_r}$) Let Γ be an ST with respect to $\mathbf{S}$ containing $\mathbf{MT_r}$. Then $\Box\varphi \in \Gamma \overset{\mathbf{T_r}}{\Rightarrow} \varphi \in \Gamma$ and $\varphi \in \Gamma \overset{\mathbf{T_r}}{\Rightarrow} \Diamond\varphi \in \Gamma$, so $\Gamma\mathcal{R}\Gamma$. Therefore $\mathcal{M}$ is reflexive.

(4_r) Let Γ, Δ, Σ be STs with respect to partial systems containing $\mathbf{M4_r}$ and suppose (1) $\Gamma\mathcal{R}\Delta$ and (2) $\Delta\mathcal{R}\Sigma$. We have to show that $\Gamma\mathcal{R}\Sigma$. This holds since $\Box\varphi \in \Gamma \overset{4}{\Rightarrow} \Box\Box\varphi \in \Gamma \overset{(1)}{\Rightarrow} \Box\varphi \in \Delta \overset{(2)}{\Rightarrow} \varphi \in \Sigma$ and $\varphi \in \Sigma \overset{(2)}{\Rightarrow} \Diamond\varphi \in \Delta \overset{(1)}{\Rightarrow} \Diamond\Diamond\varphi \in \Gamma \overset{4}{\Rightarrow} \Diamond\varphi \in \Gamma$. Therefore $\mathcal{M}$ is transitive.

(5_r) Let Γ, Δ, Σ be STs with respect to partial systems containing $\mathbf{M5_r}$ and suppose (1) $\Gamma\mathcal{R}\Delta$ and (2) $\Gamma\mathcal{R}\Sigma$. We have to show that $\Delta\mathcal{R}\Sigma$. This holds since $\Box\varphi \in \Delta \overset{(1)}{\Rightarrow} \Diamond\Box\varphi \in \Gamma \overset{5}{\Rightarrow} \Box\varphi \in \Gamma \overset{(2)}{\Rightarrow} \varphi \in \Sigma$ and $\varphi \in \Sigma \overset{(2)}{\Rightarrow} \Diamond\varphi \in \Gamma \overset{5}{\Rightarrow} \Box\Diamond\varphi \in \Gamma \overset{(1)}{\Rightarrow} \Diamond\varphi \in \Delta$. Therefore $\mathcal{M}$ is euclidean.

■

Unter dem Lindenbaum

The completeness proof given above presupposes a suitable form of the good-old Lindenbaum lemma, viz. if $\Gamma \nvdash \varphi$ then Γ can be extended to an ST (EST, FST) Σ such that $\varphi \notin \Sigma$. Although this was sufficient for the Henkin-completeness proof of the core systems, for more sophisticated applications we need a strengthened form of this lemma. Recall that

$$\Gamma \vdash \Delta \quad \text{iff} \quad \bigwedge\!\!\!\bigwedge \Gamma' \vdash \bigvee\!\!\!\bigvee \Delta' \text{ for finite } \Gamma' \subseteq \Gamma, \Delta' \subseteq \Delta$$

Proposition 4.6 (generalized Lindenbaum lemma)
If $\Gamma \nvdash \Delta$ then Γ can be extended to an ST *Σ such that $\Sigma \cap \Delta = \emptyset$. Moreover, Σ may be chosen to be a* CST *if $\Delta \neq \emptyset$, and to be an* FST *if $\Gamma \neq \emptyset$.*

Proof sketch: Let $\{\varphi_n\}_n$ be an enumeration of the modal formulas with countable repetition. Starting with $\Sigma_0 = \Gamma$, recursively define Σ_{n+1} by adding φ_n to Σ_n iff $\Sigma_n \vdash \varphi_n$. Moreover, if φ_n is a deducible disjunction, also add disjuncts ψ such that $\Sigma_n, \psi \nvdash \Delta$. Let $\Sigma = \bigcup_n \Sigma_n$. By the definition of provability from arbitrary sets and the countable repetition, Σ is an $\mathbf{M}$–theory, and the prudent clause for disjunction guarantees saturation. An easy induction using the cut theorem establishes $\Sigma_n \nvdash \Delta$ and therefore $\Sigma \nvdash \Delta$. So $\Sigma \cap \Delta = \emptyset$ and $\Gamma \subseteq \Sigma_0 \subseteq \Sigma$.

If $\Delta \neq \emptyset$ the same construction, now for $\mathbf{M}^+$, shows that Σ is consistent, for otherwise *ex falso* causes the derivation of all formulas. If $\Gamma \neq \emptyset$ then also $\Sigma \neq \emptyset$. ■

We end this subsection with some considerations concerning more general types of completeness. Is it possible to generalize theorem 4.10 such that it describes completeness of a *type* of schemata by some general condition? To make one more step in this direction, we first prove completeness with respect to a more complicated condition. To this purpose we introduce the structural condition called *confluence*[17] and the deduction rule $\mathbf{G_r}$.

confluence: $\forall s, t, t' \in S\ \exists u \in S : sRt\ \&\ sRt' \Rightarrow tRu\ \&\ t'Ru$;

[17] 'Confluence' has a bewildering variety of synonyms: 'convergency' [HC84], 'directedness' or 'the diamond property' (van Benthem) and 'incestuality' [Ch80].

($\mathbf{G_r}$) $\Diamond\Box\varphi \vdash \Box\Diamond\varphi$

Theorem 4.11 (completeness of $\mathbf{G_r}$)
$\mathbf{MG_r}$ *is sound and complete for confluent models.*

Proof: soundness is straightforward. To prove completeness it suffices to show that the canonical relation $\mathcal{R}$ is confluent. So suppose for some saturated theories Γ, Δ, Δ': $\Gamma\mathcal{R}\Delta$ and $\Gamma\mathcal{R}\Delta'$, i.e. $\Box^{-1}\Gamma \subseteq \Delta \subseteq \Diamond^{-1}\Gamma$ (1) and $\Box^{-1}\Gamma \subseteq \Delta' \subseteq \Diamond^{-1}\Gamma$ (2). Proving the existence of a Σ such that $\Delta\mathcal{R}\Sigma$ and $\Delta'\mathcal{R}\Sigma$ amounts to showing that there is an ST Σ with

$$\Box^{-1}\Delta \cup \Box^{-1}\Delta' \subseteq \Sigma \subseteq \Diamond^{-1}\Delta \cap \Diamond^{-1}\Delta'$$

This can be shown by the generalized Lindenbaum lemma. For suppose that $\Box^{-1}\Delta \cup \Box^{-1}\Delta' \vdash (\Diamond^{-1}\Delta \cap \Diamond^{-1}\Delta')^c$, then there are δ_i, δ'_j, ε_k with $\Box\delta_i \in \Delta$, $\Box\delta'_j \in \Delta'$ and $\Diamond\varepsilon_k \notin \Delta \cap \Delta'$ such that $\bigwedge_i \delta_i \wedge \bigwedge_j \delta'_j \vdash \bigvee_k \varepsilon_k$ (3). Put $\varepsilon = \bigvee_k \varepsilon_k$. Since for all k $\Diamond\varepsilon_k \notin \Delta \cap \Delta'$ and $\Delta \cap \Delta'$ is an ST, we obtain $\bigvee_k \Diamond\varepsilon_k \notin \Delta \cap \Delta'$, and so by R14: $\Diamond\varepsilon \notin \Delta \cap \Delta'$ (4). By R16 and R17, (3) $\Rightarrow$ $\Box \bigwedge_i \delta_i \wedge \Diamond \bigwedge_j \delta'_j \vdash \Diamond\varepsilon$ (5). Thus (R13) $\Box \bigwedge_i \delta_i \in \Delta$ and $\Box \bigwedge_j \delta'_j \in \Delta' \overset{(2)}{\Rightarrow} \Diamond\Box \bigwedge_j \delta'_j \in \Gamma \overset{G_r}{\Rightarrow} \Box\Diamond \bigwedge_j \delta'_j \in \Gamma \overset{(1)}{\Rightarrow} \Diamond \bigwedge_j \delta'_j \in \Delta$. By (5) therefore $\Diamond\varepsilon \in \Delta$. An analogous argument shows $\Diamond\varepsilon \in \Delta'$. In all, $\Diamond\varepsilon \in \Delta \cap \Delta'$, contradicting (4). Therefore, $\Box^{-1}\Delta \cup \Box^{-1}\Delta' \not\vdash (\Diamond^{-1}\Delta \cap \Diamond^{-1}\Delta')^c$ and the generalized Lindenbaum lemma provides the desired Σ. ∎

In fact, if **S** extends **M** by any of the principles discussed before (including $\mathbf{M^+}$ and **M***), $\mathbf{SG_r}$ is sound and complete for confluent **S**–models. We claim that $\mathbf{G_r}$ itself can also be generalized.

Conjecture 4.1
$\mathbf{G_r^{k,\ell,m,n}}$ *is sound and complete for* k, ℓ, m, n*–confluent models.*[18]

Conjecture 4.2 *If* $\vdash \varphi \rightarrow \psi$ *is sound and complete for classical Kripke models satisfying constraint C, the pair of rules* $\varphi \vdash \psi$ *and* $\tilde{\psi} \vdash \tilde{\varphi}$ *is sound and complete for partial modal models satisfying C, provided that* φ *and* ψ *are* **positive** *schemata (i.e. only use meta-variables,* $\Box, \Diamond, \wedge, \vee$*).*

The former conjecture clearly covers all the cases treated above. The latter is more general, of course. If true, it not only covers the Geachean rule schemata, but also the converse rules (such as McKinsey's), for which the frame conditions are not in general first–order definable. On the other hand, the second conjecture would imply the first-order definability of *Sahlqvist rules*, i.e. $\varphi \vdash \psi$ and its contrapositive dual $\tilde{\psi} \vdash \tilde{\varphi}$, where φ and ψ are positive schemata, and φ is constructed from $\Box^n p_i, \Diamond, \wedge, \vee$ ($n \geq 0$). The celebrated Sahlqvist theorem for normal modal logic provides first-order definability of $\vdash \varphi \rightarrow \psi$ for φ and ψ meeting the requirements above, so the second conjecture

[18] $\mathbf{G_r^{k,\ell,m,n}}$ stands for $\Diamond^k\Box^\ell\varphi \vdash \Box^m\Diamond^n\varphi + \Diamond^m\Box^n\varphi \vdash \Box^k\Diamond^\ell\varphi$, mimicking the familiar generalized Geach axiom $\mathbf{G^{k,\ell,m,n}}$ of Lemmon & Scott, cf. [Ch80] or [HC84]. k, ℓ, m, n–confluence is expressed by $\forall s, t, t'\exists u : sR^k t \;\&\; sR^m t' \Rightarrow tR^\ell u \;\&\; t'R^n u$.

would have a 'partialized Sahlqvist theorem' as a corollary.[19] For the moment, we can only speculate on the proof of a transfer result such as the second conjecture. One possible transfer technique is the Gilmore/Langholm translation procedure, adapted for modal logic.[20] We leave this to future research.

4.6.2 The finite model property

The finite model property (FMP) says that validity can be determined on finite models. The FMP provides decidability for finitely axiomatized normal modal systems. The finite model property of normal modal systems is usually established by the method of filtration. This section shows a transplant of the filtration method in the body of partial modal logic.

Filtrations

Filtration transforms an arbitrary model into a finite model. It acts by means of an identification of worlds with respect to a particular finite set of formulas. More precisely, let Γ be a finite set that is closed under subformulas, and $M = \langle S, R, V\rangle$ be a partial modal model. First define an equivalence relation $\simeq$ on S with respect to Γ and M:[21]

$$s \simeq s' \quad \text{iff} \quad (s \models \alpha \Leftrightarrow s' \models \alpha) \;\&\; (s \dashv\!\!\!\mid \alpha \Leftrightarrow s' \dashv\!\!\!\mid \alpha) \text{ for all } \alpha \in \Gamma.$$

Then S may be partitioned according to $\simeq$, forming equivalence classes $[s]_{\simeq}$ defined by $[s]_{\simeq} = \{s' \mid s \simeq s'\}$:

Definition 4.2 (filtration of model)
Let $M = \langle S, R, V\rangle$ be a partial modal model and Γ a finite set, closed under subformulas. A filtration of M through Γ is a model $M' = \langle S', R', V'\rangle$ such that

1. $S' = S/_{\simeq} = \{[s]_{\simeq} \mid s \in S\}$
2. *if* sRt *then* $[s]R'[t]$
3. *if* $[s]R'[t]$ *then* $M, s \models \Box\alpha \Rightarrow M, t \models \alpha \;\&\; M, t \dashv\!\!\!\mid \alpha \Rightarrow M, s \dashv\!\!\!\mid \Box\alpha$ *for all* $\Box\alpha \in \Gamma$;
4. $V'(p, [s]) = V(p, s)$ *if* $p \in Prop \cap \Gamma$ *(else* $V'(p, [s]) = 1$*).*

---

[19] In other words, such a Sahlqvist theorem has intermediate strength with respect to the given conjectures.

[20] Cf. chapter 2

[21] We omit the indices Γ and M of $\simeq$ when clear from context. An alternative definition is: $s \simeq s' \Leftrightarrow [\![\alpha]\!](s) = [\![\alpha]\!](s')$ for all $\alpha \in \Gamma$.

Notice that a filtered model is finite. Definition 4.2 allows of several kinds of filtrations[22], in particular with respect to accessibility, but it suffices to establish the desired preservation (restricted to Γ):

Proposition 4.7 (filtration lemma)
If M' is a filtration of M through Γ, then for all s in M and $\alpha \in \Gamma$: $M, s \models \alpha \Leftrightarrow M', [s] \models \alpha$ and $M, s \rightleftharpoons \alpha \Leftrightarrow M', [s] \rightleftharpoons \alpha$.

Proof: (by simultaneous induction on the structure of $\alpha \in \Gamma$). Notice this holds for the propositional variables in Γ by definition, and the inductive steps for the connectives are straightforward. This leaves us the modal case $\Box\alpha \in \Gamma$.

- Assume $M, s \models \Box\alpha$ and let t be any situation such that $[s]R'[t]$, then by clause *3*, $M, t \models \alpha$, and so by IH, $M', [t] \models \alpha$. Therefore $M', [s] \models \Box\alpha$.
  Next assume $M', [s] \models \Box\alpha$ and let t be any situation such that sRt, then by clause 2, $[s]R'[t]$, and so $M', [t] \models \alpha$. Thus (by IH) $M, t \models \alpha$, in all $M, s \models \Box\alpha$.
- Assuming $M, s \rightleftharpoons \Box\alpha$, there is a t with sRt and $M, t \rightleftharpoons \alpha$, thus by clause *2* $[s]R'[t]$, and by IH $M', [t] \rightleftharpoons \alpha$. Therefore $M', [s] \rightleftharpoons \Box\alpha$.
  Now assume $M', [s] \rightleftharpoons \Box\alpha$, so there is a t such that $[s]R'[t]$ and $M', [t] \rightleftharpoons \alpha$. Thus (by IH) $M, t \rightleftharpoons \alpha$, and so by clause *3*, $M, s \rightleftharpoons \Box\alpha$.

■

The generality of the filtration method may be illustrated by the observation that the proposition obviously generalizes to non-standard connectives ($\star$, $\sim$, ...).

Standard filtrations are the finest filtration $\underline{R}$, which is the smallest R' satisfying the stated requirements, and the coarsest filtration $\overline{R}$, which is the largest R' meeting the requirements. These relations can also be obtained in a direct way:[23]

$$\begin{aligned} [s]\underline{R}[t] &\Leftrightarrow s'Rt' \text{ for some } s' \in [s] \text{ and } t' \in [t] \\ [s]\overline{R}[t] &\Leftrightarrow M, s \models \Box\alpha \Rightarrow M, t \models \alpha \ \& \\ &\quad\;\; M, t \rightleftharpoons \alpha \Rightarrow M, s \rightleftharpoons \Box\alpha \text{ for all } \Box\alpha \in \Gamma. \end{aligned}$$

Notice that due to the definitions of $\simeq$ and $[\cdot]$, the finest and coarsest filtered relations $\underline{R}$ and $\overline{R}$ are well-defined. The choice of the filtered relation R' depends on the kind of model: the filtration should obey the same structural constraints as the original model. The relation $\overline{R}$ is particularly useful, although it sometimes needs to be modified further to make it fit into a class of models.

[22] The notion of filtration introduced is thus underdefined. Although this goes unnoticed in the usual textbooks on modal logic (such as [Ch80] and [HC84]), the notion is not even provably well-defined since nothing assures $[s']R'[t']$ if $s \simeq s'$, $t \simeq t'$ and $[s]R'[t]$. All the concrete filtrations in the text to follow are well-defined, of course.

[23] Cf. [BS84] and [HC84] for similar redefinitions in the case of normal modal logic.

The FMP for special systems

Before actually proving the FMP for the special systems treated earlier in this chapter, let us ruminate a bit on the notion itself. In normal modal logic the property refers to the possibility of supplying (within the correct class of frames) finite counter-models to nonvalid formulas. In partial logic validity of rules rather than of formulas is what counts. This suggests that we look for suitable finite counterexamples to non-consequences. However, this would be asking too much: $\Gamma \not\vdash \varphi$ may have no finite countermodels. The point is that Γ may be infinite, and since Γ has to be verified, this may obstruct the construction of a finite counterexample. So it seems Γ has to be restricted to a *finite* set of formulas. After all, we do not want to require this of a normal system either (**S5** would have this 'strong finite model property', but most other systems would not).

Definition 4.3 (finite model property)
A modal system **S** *has the* FMP *iff for every finite set of formulas* Σ *and every* φ*:* $\Sigma \vdash_{\mathbf{S}} \varphi \Leftrightarrow (M, s \models \Sigma \Rightarrow M, s \models \varphi)$ *for all finite* **S***-models* M.

It turns out that the systems described in theorem 4.10 (i.e. extensions of $\mathbf{M}$, $\mathbf{M}^+$ or $\mathbf{M}^*$ by a selection of rules out of $\mathbf{D_r}$, $\mathbf{T_r}$, $\mathbf{4_r}$ and $\mathbf{5_r}$) share the FMP. A problem in proving this theorem is that the correct type of filtration depends on the kind of structural condition and, moreover, these requirements cannot simply be superimposed. In other words, each system **S** that belongs to the set indicated above has to be checked separately. Fortunately, we do not have to deal with $3 \cdot 2^4$ systems since (a) the choice of the core system is unimportant for the argument and (b) there is considerable redundancy with regards to the choice of the extra rules. To wit, for a core system $\mathbf{M}^\bullet$,[24] the eleven possible systems can be put into a lattice, see figure 4.1.

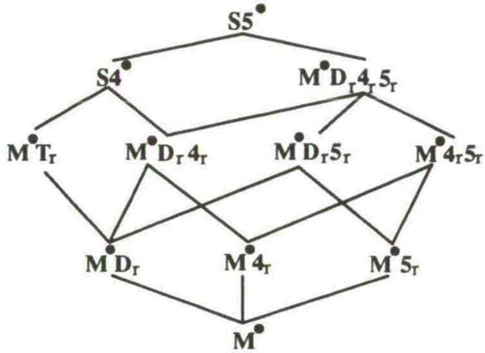
Theorem 4.12 (finite model property)
The complete systems described in theorem 4.10 all have the finite model property.

Proof: One side is trivial, since if $\Sigma \vdash_{\mathbf{S}} \varphi$, then by the completeness theorem $\Sigma \models_{\mathbf{S}} \varphi$, and so *a fortiori* $M, s \models \Sigma \Rightarrow M, s \models \varphi$ for every finite **S**-model M. For the other direction, suppose that for a finite Σ: $\Sigma \not\vdash_{\mathbf{S}} \varphi$, then by completeness there is a model M such that $M, s \models \Sigma$ and $M, s \not\models \varphi$. A suitable filtration with respect to Γ, the closure of $\Sigma \cup \{\varphi\}$ under subformulas, then produces the desired finite model M', which is also a counterexample for $\Sigma \models_{\mathbf{S}} \varphi$. So what is a suitable filtration? We need not go through 11 separate cases, since some general remarks give more reductions. First notice that (well-defined) filtrations preserve reflexivity and seriality. Reflexivity, for example, since $sRs \Rightarrow [s]R'[s]$ by clause 2 of definition 4.2 and the fact that $s \in [s]$. Next we discuss some salient cases, leaving the other ones to the reader:

- For a core system $\mathbf{M}^\bullet$ there is no structural condition on accessibilty, so *every* filtration, for example the finest one, qualifies. By the above remark, the same holds for $\mathbf{M}^\bullet\mathbf{D_r}$ and $\mathbf{M}^\bullet\mathbf{T_r}$. (This already deals with 9 systems). Since $\underline{R}$ does not preserve transitivity and euclidicity, we need other methods for dealing with $\mathbf{4_r}$ and $\mathbf{5_r}$.

[24] $\mathbf{M}^\bullet$ is either $\mathbf{M}$ or $\mathbf{M}^+$ or $\mathbf{M}^*$.

Figure 4.1: Lattice of special systems



- For $\mathbf{M^\bullet 4_r}$ the coarsest filtration comes quite close, but still does not preserve transitivity.[25] $\underline{R}$ can be augmented in the following way: let Γ be as before and[26]

$$[s]R'[t] \Leftrightarrow \begin{cases} M,s \models \Box\alpha \Rightarrow M,t \models \Box\alpha \wedge \alpha & \& \\ M,t \dashv\!\!\!\!\equiv \Box\alpha \wedge \alpha \Rightarrow M,s \dashv\!\!\!\!\equiv \Box\alpha & \text{for all } \Box\alpha \in \Gamma. \end{cases}$$

  R' is a correct filtration since: (i) it is well-defined (by bidirectionality of the definition of R' and definition of $\simeq$); (ii) clause *2* of definition 4.2 holds, since $M,s \models \Box\alpha \Rightarrow$ (R is transitive) $M,s \models \Box(\Box\alpha \wedge \alpha)$, and if sRt then $M,t \dashv\!\!\!\!\equiv \Box\alpha \wedge \alpha \Rightarrow M,s \dashv\!\!\!\!\equiv \Box\alpha$ (by transitivity); (iii) clause *3* of definition 4.2 obviously holds; (iv) transitivity is preserved, since if $[s]R'[t]$ and $[t]R'[u]$ then if $M,s \models \Box\alpha$ for some $\Box\alpha \in \Gamma$, then $M,t \models \Box\alpha$ and so $M,u \models \Box\alpha \wedge \alpha$ and *vice versa* for $\dashv\!\!\!\!\equiv$. Therefore $[s]R'[u]$.
  Since R' preserves seriality and reflexivity too, this shows the FMP for 3 positions in the lattice above (among which is $\mathbf{S4^\bullet}$). A similar 'augmented coarsest' filtration (to wit: with $\Leftrightarrow$s instead of $\Rightarrow$s in the definition of R') deals with $\mathbf{M^\bullet 4_r 5_r}$.

- For $\mathbf{M^\bullet 5_r}$ a similar adaptation of the coarsest filtration does not work[27], but the coarsest filtration itself will do, provided that Γ is extended to the infinite set Γ^* by prefixing arbitrary modalities (i.e. sequences consisting of $\neg, \Box, \Diamond$) to the formulas of Γ. Notice that infinity of Γ^* does not preclude a proper filtration (although formally definition 4.2 has to adapted). The filtration $\overline{M}$ through Γ^* will be finite, because the system allows for only 14 logically distinct modalities: the positive modalities can be reduced by $\Diamond\Diamond\Box = \Diamond\Box\Box = \Diamond\Box$, $\Box\Box\Diamond = \Box\Diamond\Diamond = \Box\Diamond$, $\Box\Diamond\Box = \Box\Box\Box = \Box\Box$, $\Diamond\Box\Diamond = \Diamond\Diamond\Diamond = \Diamond\Diamond$. $\overline{R}$ is correctly defined and now preserves euclidicity. As before this also proves the FMP for $\mathbf{M^\bullet D_r 5_r}$ and $\mathbf{M^\bullet T_r 5_r}$=$\mathbf{S5^\bullet}$.

■

[25] [BS84, p.45] wrongly claim preservation of symmetry and transitivity by the coarsest filtration; in fact any attempt along this line will fail, since **KB** and **K4** have infinitely many logically distinct modalities, cf. [Ch80, p.169], so the text approach to $\mathbf{M^\bullet 5_r}$ cannot be used here. Notice, however that symmetry is preserved by the *finest* filtration.

[26] This is basically the approach of [Ch80, p.105,106] and [HC84, p.143,144].

[27] Cf. [Ch80, p.108,109]

Apart from the intended applications on defining finite characterising models, the FMP is important since it provides decidability of the question whether $\Sigma \vdash_S \varphi$ for finite Σ and *Prop*, S being one of the systems discussed above. For S is characterized by a finite number of deduction rules, so the proofs and the deducible formulas are enumerable. With respect to *Prop* the finite models are also enumerable, and the FMP predicts that counterexamples to non-theorems must be among them. Together this provides decidability.

For classes of models with equivalence accessibility there are only finitely many essentially different finite models if *Prop* is finite as well. Hence the FMP provides a route to showing logical finiteness of a logic corresponding to such a class. In the next subsection we provide a more traditional route to this result, *via* normal forms that are useful anyway.

4.6.3 Normal forms and logical finiteness

A general normal form

In general we may put formulas of partial modal logic into a normal form which is an analogue of the usual disjunctive normal form of **pL** (DNF), which is a disjunction of conjunctions of literals (p_i or $\neg p_i$). One particular modal disjunctive normal form (MDNF) is a slight modification of an elegant and useful form suggested by Fine.[28] First we define the auxiliary notion of conjunctive form. Conjunctive forms act as modal state-descriptions. If Ψ is any set of formulas, let $(\neg)[\Psi]$ consist of the formulas in Ψ and their negations, i.e. $(\neg)[\Psi] = \Psi \cup \neg[\Psi]$.

Definition 4.4 (conjunctive forms)
F_d the set of conjunctive forms of degree d is defined recursively by:

- $F_0 = \{\bigwedge\!\!\!\bigwedge \Lambda \mid \Lambda \subseteq (\neg)[Prop]\}$;
- $F_{d+1} = \{\bigwedge\!\!\!\bigwedge \Lambda \wedge \bigwedge\!\!\!\bigwedge \Phi \mid \Lambda \subseteq (\neg)[Prop] \;\&\; \Phi \subseteq (\neg)[\Diamond[F_d]]\}$.

A general MDNF is now simply a finite disjunction of conjunctive forms. The proposition below means that formulas of depth d may be normalized to MDNFs of degree d.[29]

[28] *Vide* [Fi75b], with a somewhat different terminology: our 'conjunctive forms' are Fine's normal forms. Since Fine discusses normal systems his conjunction $\bigwedge\!\!\!\bigwedge(\neg)_i\Diamond\alpha_i$ is over *all* members of F_d, ours is over a subset of F_d. Fine uses his MDNF for a completeness proof in which the canonical worlds are conjunctive forms rather than maximal consistent sets of formulas. Jan Jaspars [Ja92] recently has found another normal form, which though more complex, enables a simple completeness proof, similar to [Fi75b].

[29] The (modal) depth of a formula is defined by: $d(p_i) = 0$, $d(\neg\varphi) = d(\varphi)$, $d(\varphi \wedge \psi) = d(\varphi \vee \psi) = \max(d(\varphi), d(\psi))$ and $d(\Box\varphi) = d(\Diamond\varphi) = d(\varphi) + 1$.

Proposition 4.8 (modal disjunctive normal form)
Every modal formula of maximum depth d is equivalent to an MDNF *of degree d.*

Proof: by induction on the depth d of a formula φ.

- if $d = 0$, φ is in pL. The DNF for **pL** still holds for the partial core system **M** (c.q. **rL**): First eliminate $\rightarrow$ and $\leftrightarrow$. By virtue of the rules of double negation and de Morgan's laws, negations can be pushed inwards until they reach the atoms. Then by the distributivity and associativity (section 3.2), the propositional DNF is obtained.
- assume the conjecture to hold for all φ of depth $\leq d$ (IH), and consider a particular φ of depth $d + 1$. As in the basic step, φ can be shown to be equivalent to a DNF where the conjuncts are literals or of the form $\Diamond\psi$ or $\neg\Diamond\psi$ with $d(\psi) \leq d$. By (IH) ψ is equivalent to a disjunction of conjunctive forms of degree d, say for some finite set of $\gamma_i \in F_d$: $\psi \dashv\vdash \bigvee_i \gamma_i$. Therefore $\Diamond\psi \dashv\vdash \bigvee_i \Diamond\gamma_i$. By the lemma below, the core system **M** is closed under substitutions of equivalents, so we may replace all such ψ by such $\bigvee_i \gamma_i$ in φ. Applying the propositional DNF once more, it follows that φ is equivalent to a disjunction of conjunctions of literals and (negations of) of diamonds of conjunctive forms of degree d, i.e. to disjunctions of conjunctive forms of degree $d + 1$.

■

Here we used an important feature of the system **M**:

Lemma 4.6 (substitution under equivalence)
Let $\gamma(\alpha)$ stand for a formula γ which has α as a sub-formula. Then $\alpha \dashv\vdash_{\mathbf{M}} \beta \Rightarrow \gamma(\alpha) \dashv\vdash_{\mathbf{M}} \gamma(\beta)$.

Proof: a straightforward induction on the structure of γ. Notice that the step of negation is licensed by the fact that contraposition holds for **M**. ■

Notice that the lemma does not hold for $\mathbf{M}^+$ and **M***. At first sight the restriction to system **M** may be thought to deprive the previous proposition of its generality, but then again, all steps performed in the proof do not go beyond **M**. So the given normal form holds for all extensions of **M**.

A special normal form

In the case of **S5**–like partial systems we can obtain a much stronger result. Recall that $\mathbf{S5}^-$ is the partial rule-based counterpart of **S5**–deduction without *(modal) ex falso* and *(modal) tertium non datur)*. For $\mathbf{S5}^-$ we will show a reduction to formulas of modal depth ≤ 1. Then the result can be put into an MDNF of degree 1, which is now a disjunction of conjunctions of formulas of the form φ, $\Box\varphi$ or $\Diamond\varphi$ where $\varphi \in$ pL. First we need some equivalences in $\mathbf{S5}^-$.

Lemma 4.7 (absorption)
In $\mathbf{S5}^-$ the following equivalences hold:

$$\Box\Box\varphi \dashv\vdash \Diamond\Box\varphi \dashv\vdash \Box\varphi \quad \Box(\Box\varphi \vee \psi) \dashv\vdash \Box\varphi \vee \Box\psi \quad \Diamond(\Box\varphi \wedge \psi) \dashv\vdash \Box\varphi \wedge \Diamond\psi$$
$$\Diamond\Diamond\varphi \dashv\vdash \Box\Diamond\varphi \dashv\vdash \Diamond\varphi \quad \Diamond(\Diamond\varphi \wedge \psi) \dashv\vdash \Diamond\varphi \wedge \Diamond\psi \quad \Box(\Diamond\varphi \vee \psi) \dashv\vdash \Diamond\varphi \vee \Box\psi$$

Proof:[30] We only show the equivalences in the top line; the dual statements in the second line are analogous. First for the left-hand equivalence:
$\Box\varphi \overset{\mathbf{T_r}}{\Rightarrow} \Diamond\Box\varphi \overset{\mathbf{5_r+L_r}}{\Rightarrow} \Box\Diamond\Box\varphi \overset{\mathbf{5_r+L_r}}{\Rightarrow} \Box\Box\varphi \overset{\mathbf{T_r}}{\Rightarrow} \Box\varphi$.
These reductions of iterations can be used to prove the middle equivalence:
$\Box(\Box\varphi \vee \psi) \overset{\mathbf{R18}}{\Rightarrow} \Diamond\Box\varphi \vee \Box\psi \overset{\mathbf{5_r,R5,6}}{\Rightarrow} \Box\varphi \vee \Box\psi \overset{\uparrow\mathbf{,R5,6}}{\Rightarrow} \Box\Box\varphi \vee \Box\psi \overset{\mathbf{L_r,R5,6}}{\Rightarrow} \Box(\Box\varphi \vee \psi)$.
And similarly for the equivalence in the right-hand corner:
$\Box(\Diamond\varphi \vee \psi) \overset{\mathbf{R18}}{\Rightarrow} \Diamond\Diamond\varphi \vee \Box\psi \overset{\uparrow\mathbf{,R5,6}}{\Rightarrow} \Diamond\varphi \vee \Box\psi \overset{\mathbf{5_r,R5,6}}{\Rightarrow} \Box\Diamond\varphi \vee \Box\psi \overset{\mathbf{L_r,R5,6}}{\Rightarrow} \Box(\Diamond\varphi \vee \psi)$. ∎

Proposition 4.9 (normal forms for partial S5 systems)
For **S5**$^-$*, every formula has a modal disjunctive normal form of degree 1.*

Proof: by induction on the depth d of a formula φ (cf. the proof of proposition 4.8). The basic case $d = 0$ is now trivial. For the induction step, assume the conjecture to hold for all φ of depth $\leq d$ (IH), and consider a particular φ of depth $d + 1$. Again start by eliminating $\rightarrow$, $\leftrightarrow$ and $\Box$, and push negations inwards until they reach $\Diamond$s or propositional variables. So φ is equivalent to a formula composed of $\wedge$, $\vee$, $(\neg)p_i$ and subformulas of the form $\Diamond\psi$ or $\neg\Diamond\psi$ with $d(\psi) \leq d$. By (IH), for $\alpha_{ij}, \beta_{ik}, \gamma_i \in$ pL:

$$\psi \dashv\vdash \bigvee_i (\alpha_i \wedge \bigwedge_j \Diamond\beta_{ij} \wedge \bigwedge_k \Box\gamma_{ik}).$$

So $\Diamond\psi \dashv\vdash \bigvee_i \Diamond(\alpha_i \wedge \bigwedge_j \Diamond\beta_{ij} \wedge \Box \bigwedge_k \gamma_{ik} \dashv\vdash \bigvee_i(\Diamond\alpha_i \wedge \bigwedge_j \Diamond\beta_{ij} \wedge \bigwedge_k \Box\gamma_{ik})$, where the last step repeatedly uses the absorption lemma. The substitution lemma entails that φ may be reduced to a depth 1 formula. This can be brought into a DNF with conjuncts of the form δ, $\Diamond\delta$, $\neg\Diamond\delta$, $\Box\delta$ or $\neg\Box\delta$ where δ is purely propositional. Finally, simply replace $\neg\Diamond$ and $\neg\Box$ by $\Box\neg$ and $\Diamond\neg$, respectively.[31] ∎

Corollary 4.2 *The extensions of* **S5**$^-$ *are logically finite.*

This particular consequence of the **S5**$^-$ normal form can, of course, be shown in a semantic way as well, but the particular normal form has advantages, for example, in constructing characterizing finite models for **S5**$^-$ knowledge.

4.7 Conclusion

Despite the additional richness of modal logic we may conclude that the results for the standard cases parallel that of propositional logic.

Again different values for semantic parameters give rise to various systems of logic.[32] The relation between semantics and logical system has been given in terms of completeness theorems. In particular, under (mixed) falsifiable validity on coherent models, we were able to give a new partial semantics for the modal system **K**.

[30] Of course, the given reductions also follow easily from the strong completeness proof and the properties of an equivalence relation. Deductive proofs are more concise, however.

[31] A slightly stronger proposition, viz. that every **S5**$^-$–formula can be reduced to a Fine MDNF of degree 1 has a similar, though somewhat more laborious proof.

[32] See section 3.5 for an overview of validity and rule concepts.

In order to arrive at the full picture of what we consider the standard options for partial modal logic, we reinspect *total* (possibly overdefined) models. *Duality* and some simple observations help to fill in some slots. By duality, truth on a total model can be transformed into non-falsity on a coherent model, and *vice versa*. So validity on total models can be reduced to validity on coherent models. So, essentially two types of models are of prime interest for partial logic: models with arbitrary situations and models with only coherent situations. For relative validity (logical consequence), both types of models provide systems which are logically interesting; the more general modal situation semantics is sometimes technically easier, whereas the more restricted coherent partial semantics often has greater appeal. By contrast, in some cases, such as verification on coherent models with *absolute* or *mixed* validity, the resulting rule systems are trivial. This can be shown analogously to our consideration for the propositional counterparts of these cases in section 3.5. The completeness results for the core logics are summarized in table 4.1.

Table 4.1: partial modal logics

	possible worlds	coherent situations	total situations	general situations
VERIF	$\mathbf{K}$	$\emptyset$	$\mathbf{K}$	$\emptyset$
VERIF_{rel}	$\mathbf{K_r}$	$\mathbf{M}^+$	$\mathbf{M}^*$	$\mathbf{M}$
VERIF_{mix}	$\mathbf{K_r}$	$\emptyset$	$\mathbf{K_r}$	$\emptyset$
FALSIF	$\mathbf{K}$	$\mathbf{K}$	$\emptyset$	$\emptyset$
FALSIF_{rel}	$\mathbf{K_r}$	$\mathbf{M}^*$	$\mathbf{M}^+$	$\mathbf{M}$
FALSIF_{mix}	$\mathbf{K_r}$	$\mathbf{K_r}$	$\emptyset$	$\emptyset$

Although we may dispense with total models, they are viable for showing the connection between syntax and semantics. The most interesting case is for relative rules, since they provide new systems. The systems in the triplet $\mathbf{M}$, $\mathbf{M}^+$ and $\mathbf{M}^*$ are shown to be complete for relative verification on the classes of modal situation models, partial Kripke models and total modal models, repectively. Similar to the propositional case, we display this by means of twin lattices.

$$\begin{tikzcd}[column sep=small] & \mathbf{K_r} \arrow[dl, dash] \arrow[dr, dash] & \\ \mathbf{M}^+ \arrow[rr, leftrightarrow, "contr."] & & \mathbf{M}^* \\ & \mathbf{M} \arrow[ul, dash] \arrow[ur, dash] & \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & \text{coh+tot} \arrow[dl, dash] \arrow[dr, dash] & \\ \text{coh} \arrow[rr, leftrightarrow, "dual."] & & \text{tot} \\ & \text{gen} \arrow[ul, dash] \arrow[ur, dash] & \end{tikzcd}$$

Recall that the lattice meet should not be construed as intersection of the generated systems, but as intersection of their finite descriptions (R1-19, etcetera). Intersection of the systems generated by $\mathbf{M}^+$ and $\mathbf{M}^*$ would lead to the modal system for double-barrelled consequence on coherent models. This type of validity seems to be modelled by the system **M** to which two principles are added, the propositional and modal rules of '*ex falso sequitur tertium non datur*':

(R8^{+*}) $\varphi \wedge \neg\varphi \vdash \psi \vee \neg\psi$

(R19^{+*}) $\Diamond(\varphi \wedge \neg\varphi) \vdash \Box(\psi \vee \neg\psi)$

Since the top element of the lattice, $\mathbf{K_r}$, is equivalent to the join of $\mathbf{M}^+$ and $\mathbf{M}^*$, and these rules are clearly in the style of natural deduction, we may, incidentally, have solved a problem noticed in [BS84, p.28]:

> *It seems fair to say that a deductive treatment congenial to modal logic is yet to be found.*

For we do not need the 'alien element' of the accessibility relation within the deductive system (as Hintikka and Kripke once proposed) or iterated modalities as Segerberg himself proposes. Moreover, the rule-based approach given here seems as flexible as normal systems are. So we do not need Prawitz's restriction to **S4** and **S5**, nor incorporation of systems by adding axioms as extra premises, as Segerberg seems to favour.[33]

Apart from the core systems we also showed completeness and decidability for the partial variants of celebrated modal systems such as **T**, **S4** and **S5**. The partial variants of **S5** even turned out to be logically finite. These results are particularly useful in the last part of this thesis. In general we can say that a remarkable amount of classical results can be transferred to partial modal logic.

Although on the dimension of situations we seem to have exhausted the possibilities (but one never knows ...), we surely have not examined every possible notion of validity and consequence, or every possible truth/falsity condition. Perhaps this is intrinsically impossible, since there seems to be no upper bound to the complexity of these notions. Nevertheless we seem to have succeeded in describing what we consider the standard cases, which should be the gist for any future extension.

[33] Cf. [BS84, p.30].

Part II

On Human Knowledge

Introduction to part II

Using the machinery developed in the first part, we now turn to the application of modal logic on knowledge as used or present in human beings.

As noticed in the general introduction, classical epistemic logic[34] attributes wrong properties to actual human knowledge. Generally speaking, a classical modal logic is simply too strong, in the sense that too many statements are counted as valid. For example,

(7) John knows that Mary works or does not work.

is supposed to be logically valid, which is evidently wrong. So the main research question is whether we can model human knowledge more adequately.

Partial models seem perfectly equipped for this enterprise. For example, there are very simple partial models in which John knows a situation that leaves the proposition 'Mary works' undefined. However, the same notion of validity under which (7) is invalid, would also reject

(8) Mary works or does not work.

yet to many people (8) is logically true. Now we can switch to another perspective of validity, viz. that of 'never false' (see chapter 4), which would indeed validate (8), but this would also validate (7). So we find ourselves placed between the devil and the deep sea. By modifying partial semantics, chapter 7 provides different ways out of this dilemma.

Another, more conservative possibility is to use classical models after all, and search for ways to weaken the logic. Sometimes this amounts to a simulation of partiality within a total semantics, sometimes the departure is more radical. Chapter 6 deals with these 'total'[35] logics for conscious belief and knowledge.

Before we turn to these more general accounts, we study one aspect of knowledge related to natural language: what is assumed to be known by the speaker when he is uttering some assertion or other? Apart from understanding the way in which information is conveyed in conversations, this special application also provides an easy access to the field of epistemic logic.

---

[34]The key reference here is [Hi62].

[35]Here in the sense of 'augmented classical', i.e. related to some form of possible world semantics, but, differently from part I, *not* overdefined.

Chapter 5

The use of knowledge and the knowledge of use

This chapter concerns the epistemic force of language utterances. Especially challenging for any proposal in this direction are the pragmatic paradoxes stemming from G. E. Moore. Our proposal deals with these and all other assertions. It is argued that the characterization can be considered as an explicitation and improvement of Grice's quality maxims. The given explanation presupposes a distinction of the levels of pragmatics and semantics, which can be argued for on independent grounds. The chapter also contains a discussion of earlier attempts to solve the problems at stake. We conclude that our solution seems superior to other proposals, except for one, to which it turns out to be technically equivalent.

5.1 Introduction: the problem

What do we mean when we utter a simple sentence such as

(9) It is raining.

Well, we obviously may mean that it is raining (right now).[1] But does it? This, of course, need not be the case. It is clear, especially from the point of view of a hearer, that the utterance of (9) does not entail that it rains (and therefore cannot mean that either). Still we are inclined to stick to the Tarskian paradigm[2]: (9) simply means that it is raining, i.e. (9) is true iff it is raining.

To avoid a contradiction here we have to assume that, generally speaking, the meaning of a sentence is different from the meaning of the utterance of that sentence. This calls for a fundamental question: what precisely is the difference between sentence

[1]In fact for a simple indicative sentence such as (9) an assertion is not the only, perhaps not even the most likely intention of the speaker. This does not influence the validity of our argument, however.

[2]At least with respect to the facts treated in this chapter a *static* approach to semantics seems possible. Incorporation of other phenomena may require a *dynamic* approach.

meaning and utterance meaning? And, can we derive one from the other by some general scheme?

In this chapter we focus on *declarative* utterances, i.e. utterances of (usually indicative) sentences to assert some proposition or other. We claim that a significant effect of uttering such a sentence is to bring about a certain change in the knowledge of the agents present in the context of utterance. As we will see, this epistemic effect is not reserved to the hearer, it also involves the speaker. In fact a great deal of the explanation of certain pragmatic paradoxes resides in the effect of uttering on the speaker herself.

In the rest of the chapter we proceed as follows. We first introduce the semantic representation language, a variant of modal logic. Then we return to the distinction between sentence meaning and utterance meaning. By a case study of the semantics of verbs such as *know* and *believe* we give other arguments for keeping both levels of description. Then we try to assess the additional epistemic force connected to uttering: is it speaker's *belief* or *knowledge* or what? This is where Moore's paradoxes come in. Then belief is argued to be too weak, and knowledge too strong, but there exists a satisfying middle course.

Although we take the pragmatic theory of Grice as a point of departure, we hope to demonstrate that the theory needs some modification in order to be able to describe and explain the facts. The resulting assessment of the propositional attitude involved in uttering is subsequently tested on a number of examples. Finally, there is a discussion of several other proposals to solve Moore's paradoxes.

5.2 Semantics vs. pragmatics

In this chapter we will present sentences and utterances in schematic form, and consequently we will, even in the examples, abstract somewhat from reality. Furthermore, p is meant to be a (possibly complex) declarative sentence.[3] For sheer convenience, we do not make a notational difference between a sentence p of natural language and an assertion p of the logical representation language.

The meaning of sentences such as

(10) Adam knows that p.

(11) Adam thinks (believes, supposes, ...) [4] that p

will be represented by, respectively,

(12) $K_a p$,

(13) $B_a p$.

[3]The reader may complete the examples below by replacing p by, for example, 'it is raining'.

[4]Unless stated otherwise, we will ignore the semantic differences between these words as well as the possible ambiguity between different senses of these words.

In these formulas the letters K and B stand for the logical counterparts of the English verbs 'to know' and 'to believe', and the index a represents the agent to whom that knowledge or supposition is ascribed, i.e. Adam in these examples. The modal operators B_a and K_a express propositional attitudes; they are one-place operators that produce a proposition when applied to another one, just like negation in elementary logic. Since we do not need quantification and only need individual constants as indices to modal operators, the logical language will be multi-modal propositional logic, i.e. the logic contains two families of operators $\{K_a\}_a$ and $\{B_a\}_a$, where each K_a and B_a functions as a modal operator $\Box$.

What do we gain by representing (10) and (11) by (12) and (13), respectively? Although a shorter notation is useful, this is not the prime motivation for using logical translations. One advantage is the explicitness of the representation as required for an account of the ambiguity in

(14) Adam knows that he thinks that p.

In one reading 'he' is anaforic (refers to Adam), so can be replaced by a in the logical translation. In another reading 'he' is deiktic (refers to extralinguistic context), which can be represented by using a free variable x:

(15) $K_a B_a p$

(16) $K_a B_x p$

More important, however, is the fact that, by means of this translation, the meaning of the matrix verbs in (10) and (11) has been made (more) precise. The concise, explicit and precise nature of these logical operators enables us to reason with the semantic representations and abstract from the idiosyncratic variation in the meaning of epistemic verbs. Some people use 'know' in cases where others would prefer 'believe'. Yet we only interpret words by their *standardized literal* meaning. Perhaps it is clear now that an immediate translation of sentences into formulas which reflect the speaker's intention is almost impossible, at least if we want to keep the translation function compositional and fully general. Especially relevant to what follows are the first person variants of the examples (10) and (11):

(17) I know that p.

(18) I believe that p.

Analogous to (12) and (13), we will represent the meanings of these sentences by respectively

(19) $K_i p$

(20) $B_i p$

Instead of the individual constant a we encounter a special individual variable i (representing 'I', the speaker), which may be replaced by a constant in a given context, but possibly by different constants in different contexts. Similarly for the special variable j which stands for 'you', the hearer. This treatment of deiktic elements such as 'I', which refer to objects in extralinguistic contexts, is a bit simplistic, yet it suffices for our purpose.

Now what is the effect of uttering (17)? Since saying something is in the first place related to the speaker's state of mind, the utterance meaning seems to be roughly described by

(21) $B_i K_i p$,

for people can utter (17) if they sincerely think they know that p, whereas in fact p is not the case. (19) would assign too strong an opinion to the speaker, and (20) too weak an opinion, therefore (21) seems the correct compromise between these simpler options. This is, in a nutshell, the gist of our later general proposal.

Meanwhile we have introduced a useful distinction between semantics and pragmatics. The use of this distinction becomes especially evident with respect to sentences such as (17) and (18). Without the distinction the meaning of (17) would simply be (21), for then this is the only meaning, overruling a purely semantic interpretation. This state of affairs seems to be excluded by the so-called 'truth axiom' for knowledge, which states that knowledge should be true:

(T) $K_\alpha \varphi \rightarrow \varphi$

The validity of **T** for knowledge is being disputed. A typical counter-argument runs as follows: "Consider an example such as (17). One can easily *say* φ without φ being true: people can be mistaken. So, 'to know' does not imply that it has to be the case." Those advocating this point of view insufficiently distinguish between pragmatics and semantics in the narrow sense. In this view the meaning of a sentence can differ from one context to another, and is completely determined by that context. In short: meaning and use would coincide. The complications of such a point of view are enormous.

Firstly, the distinction between semantics and pragmatics has been introduced for good reasons, of course. If 'It is hot in here' in a certain context means that, according to the speaker, a window has to be opened, and in another that an unfortunate statement has been made, then these different utterance meanings have to be explained without reference to a formal meaning of the sentence. We claim that it will be very hard to do this; at least it will be unnecessarily laborious to give a direct pragmatic interpretation in a principled, compositional way, since the various pragmatic effects of words and phrases have to be taken into account, leading to a proliferation of different readings, all except one of which usually being absent in the context of utterance.

Secondly, and more positively, there are clear intuitions that an independent, 'context-free' meaning can be attributed to a sentence. Subsequently this literal semantic meaning is the kernel on which pragmatic rules operate, or, to put it differently, form an argument for a communicative function.

Another view combines acceptance of a level of formal semantics with rejection of **T**. This point of view also produces considerable problems:

- First and above all: it conflicts with our intuition (and with what is usually accepted in epistemic logic or philosophy).
- It is unclear how the difference between 'to know' and 'to believe' (or: 'to suspect') can be accounted for in this way. At least we have deprived ourselves of the possibility to mark the difference between knowledge and belief: in our view, the **T** axiom holds for knowledge, but not for belief.
- Regardless of a possible modal distinction between K and B there is a problem of circularity: if a formal meaning of 'know' is 'suppose to know', then (one reading of) 'to know' would be synonymous with 'suppose to suppose to know', etcetera. However, to us it seems that iterating a doxastic modality (such as 'believe' or 'suppose') leads to ever weaker beliefs. Consequently, in our logic

$$B_\alpha K_\alpha \varphi \leftrightarrow B_\alpha B_\alpha K_\alpha \varphi$$

does not hold: the implication to the right is valid, but the one to the left is not.

So, if we represent (17) formally by (21), we would have mixed up semantic and pragmatic meaning. For (17) basically means (19); if p is not true, then neither is sentence (17). When person i *utters* the sentence, (17) can receive an additional value. Apart from this, such meaning-in-use already arises when a person only *thinks* of a proposition: the act of uttering is not essential here. So the aim of this chapter is primarily to formulate a general rule which ascribes the right pragmatic meaning (21) to a sentence such as (17).

5.3 Logical preliminaries

The logical syntax contains the usual connectives $\neg$ ('not'), $\wedge$ ('and'), $\vee$ ('or'), $\rightarrow$ ('if ... then') and $\leftrightarrow$ ('if and only if'), as well as the modal operators K_a ('a knows') and B_a ('a believes') for each agent a. The dual operators $\hat{K}_a$ and $\hat{B}_a$ abbreviate $\neg K_a \neg$ and $\neg B_a \neg$, respectively. Apart from individual constants ($a, b, \ldots$), we will use arbitrary individual variables ($x, y, \ldots$), and designated variables i, j.

The deductive system is basically the one proposed in [Hi62], which is still classical. Each K_a behaves according to the modal system **S4=NKT4**, and each B_a to **NKD4**, and K_a is logically stronger than B_a. Recall the following axioms and rules from chapter 4: ($\Box$ is either K_a or B_a)

(pL) all the (modal instantiations of) propositional axioms;

(K) $\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$;

(4) $\vdash \Box\varphi \rightarrow \Box\Box\varphi$;

(N) if $\vdash \varphi$ then $\vdash \Box\varphi$;

(MP) if $\vdash \varphi$ and $\vdash \varphi \to \psi$ then $\vdash \psi$;

(T) $\vdash K_a\varphi \to \varphi$;

(D) $\vdash B_a\varphi \to \hat{B}_a\varphi$;

(X) $\vdash K_a\varphi \to B_a\varphi$.

To avoid at least one type of so-called logical omniscience, **N** can be eliminated in favour of the weaker rule **I**.

(I) if $\vdash \varphi \to \psi$ then $\vdash \Box\varphi \to \Box\psi$;

We mention but one useful property that is provable in systems containing **IK**:[5]

(C!) $\vdash \Box(\varphi \wedge \psi) \leftrightarrow (\Box\varphi \wedge \Box\psi)$

A sound and complete model theory for system **IK** can be given along the lines of [Kr65a], or within *neighbourhood* semantics[6].

A further weakening of the required modal system is possible when **K** and in fact all axioms are replaced by deduction rules. A partial semantics is then obtained in the fashion of part I. Interestingly, both the elimination of **N** and the further weakening in partial modal logic does not interfere with the logical derivations to follow. So, it is possible to keep the deductive explanations in a more realistic logic of knowledge and belief, supported by an appropriate partial semantics.[7]

5.4 The epistemic force of declarative utterances

Although the 'cognitive' meaning (21) of (17) can also arise without actual utterance, with respect to communication it is necessary and essential that (17) is uttered. According to an objective and omnipresent observer ('God'), person i could indeed be in an epistemic state where (21) is valid, but i's interlocutor j cannot inspect i's mind, so this inference may not be available to j if (17) has not been uttered.

Now what is the precise connection between uttering sentence (17) and pragmatic meaning (21)? It seems as if we merely prefixed B_i to the semantic meaning (19). Generalizing this to arbitrary declarative sentences φ, one obtains what we will call the *doxastic rule*:

DOX $\quad x : `\varphi' \to B_x\varphi$.

[5]This is shown in chapter 7.

[6]See section 6.6.

[7]We return to this issue in section 5.7.

Rule DOX implies that if speaker x utters assertion φ, then, under normal circumstances, x believes that φ is true.[8] Here x : 'φ' abbreviates normal utterance of (a sentence corresponding to) φ by x, where *normal* refers to honest and sound language use. So, for example, ironical or deranged use of φ is excluded.[9] Other types of sentences, such as interrogatives and imperatives, also fall outside the scope of rule DOX. We will test the correctness of rule DOX in a number of examples.

5.4.1 Moore's paradoxes

On several occasions the British moral philosopher G. E. Moore has pointed at a puzzling problem involving self-belief (e.g. in [Mo12]). Sentences of the types

(22) p, but I do not believe that p.

(23) p, but I believe that not p.

are evidently paradoxical, for they seem to be both absurd and logically possible.

On the one hand the meaning of, for example, (23) is consistent: it represents the very common case of incorrect belief. This meaning is also satisfiable in the type of models, as will be discussed below. Yet, on the other hand, ordinary sane people, unlike philosophers and linguists, would not dream of making such an "assertion": using sentence (22) probably deprives the audience of the possibility of taking the speaker seriously.

Interestingly, even if the agent is an inveterate liar, we do not expect a remark of the above types.[10] For a persistent liar to say (22) may imply $\neg p \wedge B_i p$, which is of the form (23); and, *vice versa*, saying (23) may imply $\neg p \wedge \neg B_i \neg p$, which is of the form (22). More importantly, one *cannot* lie by saying (23), since the hearer cannot possibly believe that the speaker believes what he is saying, which seems essential in the act of lying. So, unlike what Moore apparently thought, lying does not have much influence on Moore's examples. Only in some special cases can (22) and (23) be correctly used:

- when the words do not have their common meaning, usually encoded in the intonation of the sentence, e.g.:

  - 'p, but *I* do not believe it (i.e. that p).'
    where p, for example, reflects the official opinion of your company in certain matters. In some cases a shift of meaning of words can be obtained without special emphasis, such as in Max Black's example 'Oysters are edible, but I do not think so', provided the English allows for the shift of

[8] A more accurate statement of DOX would be that every occurrence of i in the utterance by x of φ is replaced by x in the believed assertion, i.e. x : '$\varphi(i)$' $\rightarrow B_x\varphi(x)$.

[9] All subsequent 'pragmatic transformations' therefore have to be read *modulo* the assumption of normal language use.

[10] [Bl52, pp.49,50] does not seem to be fully aware of this point, although he corrects Moore, who considers lying merely 'vastly exceptional'.

meaning of 'edible' from 'apt to be eaten without getting ill' to 'food that I fancy'.

- 'p, but I do not *believe* that p.'
  where a special sense of 'belief', such as in religion or when 'believe' means *accept*, is intended. In this and the previous case the speaker in fact does not believe that p.
- 'p, but I do not *believe* that p.'
  can also be used contrastively (add: ', I *know* it'), or ironically, for example when echoing an earlier, but incorrect belief attribution to the speaker. Also, the speaker can comment on his own belief, just recognized to be false. In all these cases the speaker in fact believes that p.

• when the words do not have any meaning, such as in some forms of poetry or in linguistic examples ...

As we noticed before, sentence such as (22) are consistent, for we can easily imagine situations where the statements are verified: (22) when p is a true fact which does not belong to your belief, (23) when you incorrectly believe p. So the meanings of (22) and (23) seem to be properly represented by

(24) $p \wedge \neg B_i p$

(25) $p \wedge B_i \neg p$,

respectively.[11]

Then what gives (22) such a funny ring? An explanation of this puzzle can be supplied by DOX: the *sentence* (22) is not contradictory, but the *utterance* is! For, when (22) is uttered, it receives the pragmatic meaning

(26) $B_i(p \wedge \neg B_i p)$

which is indeed a *contradiction*, in the literal as well as the logical sense of the word. Here is a sample derivation of the (logical) contradiction:

1.	$B_i(p \wedge \neg B_i p)$	[given]
2.	$B_i p \wedge B_i \neg B_i p$	[1, **C!**]
3.	$B_i B_i p \wedge \neg B_i B_i p$	[2, **4**, **D**, **pL**]

In short, the results for these examples are in accordance with DOX. Besides, Hintikka rightly observes that if we replaced 'believe' in (22) by 'believed', no pragmatic contradiction as in (26) would arise. We confine ourselves here to the observation that incorporation of time dependencies can be solved in different ways, but that the difference between (22) and its past counterpart is anyhow explainable from the simple observation that the utterance is always linked to the moment of speaking, whereas the sentence may not; this results in different doxastic modalities and no contradiction will arise.

[11] Sometimes (22) is interpreted as (25), cf. the discussion on page 132. This does not interfere with our argument, however.

5.4.2 Epistemic counterparts of Moore's paradox

Another test-case, also analyzed by Hintikka, is

(27) I know the following: *p*, but I do not know whether *p*.

In [Hi62, p.79] (27) gets the logical translation:

(28) $K_i(p \wedge \neg K_i p \wedge \neg K_i \neg p)$

This representation explains why (27) is deviant: already the semantic contents (28) is inconsistent (by **C!** and **T**), and so is the pragmatic meaning predicted by DOX:

(29) $B_i K_i(p \wedge \neg K_i p \wedge \neg K_i \neg p)$

(the contradiction follows from that of (28) by **I** and **D**). Apparently different from Hintikka, we conclude at this point that a revision or alternative to DOX is not required to explain the anomaly in (27).

Hintikka rightly observes that the embedded sentence in (27)

(30) *p*, but I do not know whether *p*.

is indeed problematic. For, even after DOX has been applied to the semantic representation of (30), no contradiction arises.[12] Yet (30) is a strange sentence to utter (or to think, for that matter). Hintikka explains the anomaly of (30) by relating it to (27). Consider the following rule.

EPI $\quad x : `\varphi' \rightarrow K_x\varphi$

(here x : 'φ' is subjected to the same condition of normality of the utterance as in DOX.)

To avoid a dilemma of choice between DOX and EPI,[13] only one of the pragmatic rules can be maintained, and because of (30) that should not be DOX but EPI. However, this move is also wrong. If EPI were obligatory, it would imply that everything we are saying is true[14] — in other words, every honest opinion should be necessarily true, which is absurd. Something is not true merely by being asserted. What does seem to hold, however, is that the speaker has to be convinced of this truth.

The conclusion we can draw at this point is: neither only DOX, nor only EPI provides a satisfactory account of the facts. Is there another possibility left? Indeed there is: one rule different from both DOX and EPI and with, roughly speaking, intermediate strength[15], which can still account for the facts (Moore's paradoxes). Instead of immediately proposing the rule we have in mind, we will relate it to a general and well-known pragmatic theory, thus giving it independent motivation.

---

[12] The consistency of $B_i(p \wedge \neg K_i p \wedge \neg K_i \neg p)$ can easily be shown by means of a (partial) Kripke model.

[13] This dilemma occurs in Hintikka's approach, see section 5.5.

[14] Of course, we do *assume* that what we say is true; this, however, does not guarantee objective truth.

[15] Notice that it is intuitively correct that the rule is stronger than DOX, but although EPI is too strong itself, the new rule need not be implied by EPI.

5.4.3 Grice revisited

As a part of his theory of conversational implicature Grice suggested in [Gr75] the pragmatic maxim of *quality*, which is a 'Kantian' category of the more general *cooperation principle*:

Quality *Try to make your contribution one that is true*

Two special maxims are subordinated to the quality principle, viz.:

Belief *Do not say what you believe to be false.*

Evidence *Do not say that for which you lack adequate evidence.*

In our framework, the maxim of belief amounts to:

(31) $\neg(x : `\varphi' \wedge B_x \neg\varphi)$

which is equivalent to

(32) $x : `\varphi' \rightarrow \hat{B}_x \varphi$

Perhaps this reformulation makes clear that DOX is stronger than the Gricean maxim of belief. We can ask ourselves, then, whether Grice may have meant:[16]

Strengthened belief *Do not say what you do not believe to be true.*

which corresponds to DOX after all. And even if this was not Grice's intention, we prefer it over the original maxim of belief. It is true, of course, that 'I do not believe that p' is often understood as 'I believe that not p', and by contraposition and the **D** axiom, the strenghtened form of the belief maxim would be equivalent to the original one, after all. There are two severe problems for such a 'cooperative' interpretation of Grice's belief maxim: (i) unlike colloquial language, linguistic rules have to be fully explicit , and (ii) we do not see how to derive 'believing not' from 'not believing' by the Gricean maxims.

In order to represent the meaning of the Gricean maxim of evidence within our framework, we have to introduce a new operator. Let E_x stand for 'X has sufficient evidence for'. Then the maxim of evidence may be recast as

(33) $\neg(x : `\varphi' \wedge \neg E_x \varphi)$

which is equivalent to

(34) $x : `\varphi' \rightarrow E_x \varphi$

[16]The difference may be caused by the fact that Grice intended his maxims to be fully general, whereas we focus on *indicative* utterances.

What do we understand by 'sufficient evidence' in the maxim of evidence? People differ according to what they accept as adequate evidence. The sources of evidence may also vary: of course, direct perception counts as providing sufficient evidence, but a mathematical proof, a careful scientific experiment or information from a textbook may do just as well. Anyway, we claim that the propositional attitude involved in making a statement is more subjective than is suggested by 'sufficient evidence': usually people say this or that if they sincerely *believe* they are licensed to. So, what is at stake is a modification of the maxim of evidence:

Modified Evidence *Do not say that for which you do not believe to have adequate evidence.*

which can be reformulated in logic as:

(35) $x : `\varphi' \rightarrow B_x E_x \varphi.$

Combining our adaptations of the Gricean maxims of belief and evidence results in:

(36) $x : `\varphi' \rightarrow (B_x \varphi \wedge B_x E_x \varphi).$

We can simplify this pragmatic rule, if we take 'sufficient evidence' to correspond to 'justification', and identify 'knowledge' as 'true justified belief' (a familiar equation in epistemology), i.e.

(37) $K_x \varphi \leftrightarrow (\varphi \wedge B_x \varphi \wedge E_x \varphi).$

A simple deduction then shows that $B_x \varphi \wedge B_x E_x \varphi$, the right-hand side of (36), is equivalent to $B_x K_x \varphi$, given the identification (37).

$$
\begin{array}{l}
B_x \varphi \wedge B_x E_x \varphi \overset{4}{\Leftrightarrow} \\
B_x \varphi \wedge B_x B_x \varphi \wedge B_x E_x \varphi \overset{C!}{\Leftrightarrow} \\
B_x(\varphi \wedge B_x \varphi \wedge E_x \varphi) \Leftrightarrow \\
B_x K_x \varphi
\end{array}
$$

So, the joined effect of (36) and (37) is that when x says 'φ', she signifies that she believes to know φ. The resulting scheme is dubbed UTT, the (epistemic/doxastic) utterance rule.

UTT $\quad x : `\varphi' \rightarrow B_x K_x \varphi.$

We have seen that this rule follows from a reinterpretation and modification of Grice's work on conversational maxims. But does it meet the criteria mentioned earlier?

First, is UTT in between DOX and EPI? Indeed, it is, provided that K_x is positively introspective:

$$K_x\varphi \overset{4}{\Rightarrow} K_xK_x\varphi \overset{X}{\Rightarrow} B_xK_x\varphi \overset{T,I}{\Rightarrow} B_x\varphi$$

By some simple Kripke models one can show that the ordering is strict, i.e. the logical implications cannot be reversed. Recall from footnote 15 that, intuitively speaking, DOX should be subordinated to UTT, since the former rule is correct, but just not strong enough. EPI, on the contrary, is too strong, but does not really have to imply UTT. This means that the positive introspection property of knowledge may be avoided, if one chooses to.

Second, is UTT empirically adequate? Since it is stronger than DOX it automatically accounts for Moore's doxastic paradoxes. Now how about the epistemic variants? Let us reinspect (30).

(30) p, but I do not know whether p.

Application of UTT to the semantic representation of (30) produces the formula

(29) $B_iK_i(p \wedge \neg K_ip \wedge \neg K_i\neg p)$,

which was already shown to be contradictory.

In order to check proposal UTT, we will review a number of Hintikka's examples.

5.4.4 Checking the proposal

Consider the utterance of

(38) He knows that p, but I do not know it.

Hintikka[17] claims that (38) is 'sometimes (not always) a somewhat strange thing to say'. We do not share Hintikka's intuition on this point. Unfortunately, Hintikka does not motivate his restriction to 'sometimes strange to say'. As before, we maintain that (38) can only be used felicitously when one of the occurrences of 'know' is given an ironic intonation. Moreover, (38) seems ambiguous, depending on the reference of *it*. Hintikka only treats the first reading.

(39) $K_xp \wedge \neg K_ip$

(40) $K_xp \wedge \neg K_iK_xp$

That (38) is anyway pragmatically anomalous can be acounted for by the fact that UTT produces an inconsistency on both readings.[18] After applying UTT and the **T** axiom to the first reading (which by **I** also applies within a modal context), both resulting formulas are of the form $B_iK_i(\varphi \wedge \neg K_i\varphi)$. This scheme is contradictory, as can be easily shown by using the principles **C!**, **T**, **I** and **D**, similar to the argument for (29).

By contrast, as Hintikka observes, the following sentence is absolutely unproblematic:

[17] [Hi62, p.80]

[18] We omitted a similar account of 'He knows that p but I don't believe it.'

(41) He knows whether p, although I do not.

This sentence may show a threefold ambiguity, with some preference for the second reading:[19]

(42) $(K_x p \vee K_x \neg p) \wedge \neg K_i p$

(43) $(K_x p \vee K_x \neg p) \wedge \neg(K_i p \vee K_i \neg p)$

(44) $(K_x p \vee K_x \neg p) \wedge \neg K_i(K_x p \vee K_x \neg p)$

Here the utterance rule operates as a pragmatic filter: the semantic meaning represented in (44) leads to a logical inconsistency after application of UTT.[20]

Some rather simple epistemic variants of Moore's paradox have not been treated so far. As a matter of fact we merely dealt with *know whether* constructions; we now turn to *know that* cases. The reason for this postponement is a slight complication caused by the factivity of the verb *to know*. The epistemic counterpart of (23), viz.

(45) p, but I know that not p.

is, of course, already contradictory on the semantic level. A simple test to show this is to substitute a third person pronoun or a proper name for 'I' in (45): the contradiction will remain. The formal account of this contradiction hinges on the inconsistency derived by the **T** axiom of knowledge and the semantic representation $p \wedge K_i \neg p$ of (45).

The counterpart of (22), viz.

(46) p, but I do not know that p.

is again anomalous, but now the contradiction arises on the pragmatic level: unlike (46),

(47) p, but he does not know that p.

is pragmatically sound (and in accordance with our rule UTT). Still, (47) seems rather clumsy or redundant. In effect, we will only be inclined to use (47) when we want to emphasize p (or, *he*, for that matter). The reason for this is probably the factivity of the verb *to know*: 'he does not know that p' already implies p. Therefore (47) is semantically equivalent to

(48) He does not know that p.

[19]It is not clear whether the first reading really exists.

[20]This is not claimed to be the only or the best explanation of the markedness of reading (44). An alternative explanation may be based on a syntactic or semantic constraint forbidding the antecedent 'know' to be copied twice into the succedent. However, this very restriction would not account for a similar phenomenon occurring with a minor variant of the text sentence: 'He knows whether p, although I don't know it.' In this case, there seems to be but one pragmatically sound reading.

the meaning of which can be represented by:

(49) $p \wedge \neg K_x p$.

Then rule UTT explains why (47) can be asserted by an agent a, whereas

(50) I do not know that p.

with the analogous analysis

(51) $p \wedge \neg K_i p$

is ruled out by application of UTT: $B_a K_a(p \wedge \neg K_a p)$ is easily shown to be contradictory.

We may conclude that there is a reinterpretation of Grice's quality maxims from which the pragmatic rule UTT follows. The new rule seems to cover the intuitions concerning the reliability of assertions and accounts for a number of problematic cases, the so-called Moore paradox and variants thereof.

5.5 Discussion and comparison

Grice

There are some differences between our logical reconstruction of the Gricean maxims and the original cooperation principles. In particular, Grice might have objected that our reformulations (such as DOX) are genuine implications instead of implicatures which have defeasible effects. Although we have restricted the application of formulas like DOX by the built-in condition of normality[21], this may still be insufficient for pragmaticists working along the lines of Grice. Yet,

- Grice himself noticed that *quality* has a somewhat different status than the other maxims:[22]

> Indeed, it might be felt that the importance of at least the first maxim of Quality is such that it should not be included in a scheme of the kind we are constructing; other maxims come into operation only on the assumption that this maxim of Quality is satisfied.

- In fact, Grice advocates a different explanation of Moore's paradox, one that goes beyond implicatures:[23]

[21] Cf. footnote 9.
[22] [Gr75, p.46]
[23] [Gr78, p.114]

> On our account, it will not be true that when I say that p, I conversationally implicate that I believe that p; [...] it is not a natural use of language to describe one who has said that p as having, for example, "implied", "indicated", or "suggested" that he believes that p; the natural thing to say is that he has expressed (or at least purported to express) the belief that p.

Grice favours an account in which the connection between utterance and belief is more tight. Here is the procedure he had in mind which links the indicative mood to speaker's belief:[24] (S is an indicative sentence, and σ the infinitive form of S)

P1 *To utter [S] if for some B, A wants/intends B to think A thinks* σ

Rephrased in our terms this amounts to a boulomaic/doxastic scheme:[25]

(52) $i : `\varphi' \rightarrow W_i B_j B_i \varphi.$

Notice this proposal works all right for Moore's original paradox by an argument similar to the one based on DOX, because both wanting and believing should be consistent.[26]. However, Grice's procedure does not deal with the epistemic variants of Moore's paradox, and is therefore empirically incorrect.[27] In [Ha77], Harnish points at another drawback of Grice's alleged solution: the proposed explanation hinges on the positive introspection property **4** of belief. Harnish's intuition is that a Moore sentence such as (22) is strange, even if belief is not introspective, i.e. even when (26) is consistent. We will return to this interesting point in section 5.7.[28]

Gazdar

Gazdar, in [Ga79], also uses the identity of 'knowledge' and 'true, justified belief', in order to give a reinterpretation of Grice's maxim of quality. However, his conclusion is different from ours: with him, *quality* takes the form of EPI[29]. Unfortunately the intermediate steps of this replacement are not given. The superficial connection of EPI with *quality* and its submaxims of belief and evidence consists of noticing the keywords 'true', 'belief' and 'adequate proof' in these maxims. Gazdar disregards the fact that in Grice's formulation of *quality* not the truth of the proposition is demanded, but the striving for that truth.[30]

---

[24] [Gr68, p.65(ed. 1971)]

[25] W_i stands for 'I want'. Again we have taken the liberty to change Grice's 'if' in 'only if', because the converse of (52) would be unacceptable.

[26] In [KT89] we have investigated the weak modal properties of *want*.

[27] There is a stronger interpretation of P1 which does the job; we will return to this.

[28] Unfortunately, Harnish relates failure of **4** to the existence of prejudices. This is a rather dubious move, since $B_i B_i p$ should not be interpreted as 'I explicitly believe that p.

[29] See [Ga79, p. 45–48]

[30] Gazdar's statement that *quality* boils down to our EPI is incompatible with the rejection of a 'sincerity rule' by G. Lakoff, which is similar to EPI. *Vide* [Ga79, p. 32].

Gazdar does not discuss the problem that the original formulation of the maxim of belief is insufficient to cover the intuition and account for the facts. Furthermore we have found that (35) seems a better interpretation of the intention of the maxim of evidence than (34), or the original maxim. Of course the most important objection against his conception of *quality* is that EPI is, logically speaking, too strong. Gazdar does give a rather curious quotation from a lecture of Sacks, which, according to Gazdar, shows that *quality* concerns knowledge, although something quite different appears from the quotation.[31]

> That *knowing* and not, say, believing, assuming, or something else is what is involved is evidenced in the following:
>
> > "We have in the data, 'Oh she knows you're crazy hehh!' where that might be different from 'She thinks you're crazy', where the problem, I suppose, is that whatever is correct to say about what she figures, then if I say 'She knows you're crazy', it's hard for you to be in a position to say 'No, she *thinks* I'm crazy. She happens to be right.' That is to say, if some facts are assertably so, then, that somebody thinks that they're so can apparently be used in such a fashion as to say that they *know* that it's so; whether or not their thoughts turn out to have a correct basis for that result."

The rather cryptical quotation of Sacks may amount to: 'knowledge' is true belief, instead of true justified belief. Although the former equation is extremely doubtful[32], we may accept it for the sake of the argument. This still does not make Gazdar's introductory sentence comprehensible! To provide the intended reduction, we need at least two more assumptions: the quality maxim has to ensure *truth* of the utterance, i.e.

$$x : `\varphi' \rightarrow \varphi$$

and the belief maxim has to be formalized as DOX. The first assumption is obviously wrong and does not correspond to *quality*, as we have seen. The second assumption does not correspond to Grice's belief maxim either, but captures the right intention, as we have also seen. Finally, the evidence maxim seems to have been ignored in this argument.

We may conclude that extra evidence for Gazdar's EPI interpretation of *quality* can not be derived from his arguments.

[31] [Ga79, footnote 11, p. 46]. We would like to thank Kees van Deemter for helping to decode the quote.

[32] We may *believe* our favourite soccer team to win next Sunday. Now suppose it does in fact win. Can we *know* this beforehand? We do not think so, in general.

Levinson

By contrast, Levinson in his textbook on pragmatics, seems to favour a DOX interpretation of *quality* to solve Moore's paradox.[33] Again it remains mysterious how the belief maxim can be interpreted as DOX.[34] Notice that a straightforward (32) interpretation of the belief maxim

(32) $x : `\varphi' \rightarrow \hat{B}_x\varphi$

cannot account for Moore's paradoxes: both $\hat{B}_i(p \wedge \neg B_i p)$ and $\hat{B}_i(p \wedge B_i \neg p)$ are consistent, as can easily be shown by simple, verifying Kripke models. Now, even for DOX, Levinson's solution is not satisfactory, since there is no trace of a formal derivation of the intended pragmatic contradiction. Moreover, the epistemic variants of Moore's paradox and the problems they cause for an explanation along the lines of DOX, are completely absent.

Hintikka

We frequently referred to [Hi62], since it presents most relevant facts without much philosophical bias. Some of the older accounts of Moore's paradox (in particular [Mo12] and [Bl52]) are aptly discussed by Hintikka.[35]

[Hi62, p.71] proposes a rule corresponding to DOX, which, however, should be considered a definition of the notion *doxastically defensible*, rather than an empirical law. Similarly, EPI is triggered by the notion *epistemically defensible*. So, in Hintikka's terms, (30) is doxastically defensible, but epistemically indefensible.

We have avoided the cautious approach of Hintikka – our prime interest here is to propose empirically correct rules, not to construct a methodological apparatus. Moreover, in [Hi62] EPI does not replace DOX, but seems to constitute an alternative to the latter. Unfortunately, the relationship between doxastic and epistemic defensibility is quite unclear. One possible interpretation is that the hearer has the freedom to choose between DOX and EPI. This interpretation causes problems, as will be demonstrated below. Another interpretation is that both criteria are used, where failure of DOX is more serious than that of EPI. The latter interpretation is supported by Hintikka's remark that (30) is doxastically defensible, yet epistemically indefensible. However, that (30) is merely 'sometimes somewhat peculiar' and (22) 'absurd to utter'[36] appeals to a spurious difference: to us both are pragmatically 'out'. That (30) may sound better than (22) has a very simple explanation: the former has a contrastive reading 'I believe *p* but do not know it', which is pragmatically consistent, whereas a similar

[33] [Le83, p.105]

[34] Perhaps some confusion concerning the concept of modal duality has caused this nonfactual interpretation: [Le83, p.135] claims that 'not knowing whether' is the logical dual of 'knowing that', which it is not. But then again, on [Le83, p.140] he aptly demonstrates the similar logical independence of 'maybe' and 'maybe not'.

[35] See, for example, [Le78, pp.84,85] for more recent literature

[36] [Hi62, p.9]

rescue for (22) fails since 'I know p but do not believe it' is semantically contradictory. Moreover, EPI is just one way to specify what Hintikka could have meant by: [37]

> When somebody makes a statement — say utters the sentence q — we are normally led to expect that he can conceivably know that what he is saying is true or that he is at least not depriving himself of this possibility by the very form of words he is using.

The expectation to which the last quotation refers not only introduces undesirable vagueness, but even vitiates the explanation to a certain extent. For one of the normal expectations of a language user is that a speaker is consistent or at least tries to be so. Yet, in the case of (30) the latter maxim will be violated, exactly when the former expectation, namely that the speaker knows what he is saying, is maintained. In short, if the hearer is allowed to make a choice between DOX and EPI, he will be forced by the condition of consistency to choose DOX, and consequently the deviation of (30) will remain unexplained.[38]

So it seems fair to say that Hintikka's two notions 'epistemically defensible' and 'doxastically defensible' together account for Moore's paradoxes, but it is not clear how to choose between them in what circumstances. Consequently, we see no way in which such a conceptual distinction can be implemented in a compositional semantic/pragmatic interpretation: his logical representations are 'miraculous translations'.

Åqvist

Åqvist, in [Åq64], also provides an analysis of Moore's paradox by deriving a contradiction in a multi-modal logic. This is where the resemblance stops. As usual Åqvist invokes a multitude of operators, not related to different agents, but to different senses of belief: ranging from *assertion* over *belief* (in the strict sense) to *conjecture*.[39] $A_i\varphi$ ('I assert that φ') is supposed to imply $B_i\varphi$. Then Moore's paradox is resolved as follows: the sentence

(23) p, but I believe that not p.

is still translated into $p \wedge B_i\neg p$. Uttering (23) logically amounts to

(53) $A_ip \wedge B_i\neg p$,

from which the deductive system yields a contradiction. However, Åqvist's solution contains a number of oddities:

[37] [Hi62, p.78]; we treat other interpretations of this quotation in section 5.7.

[38] So a non-monotonic (default) approach will also fail, for virtually the same reason.

[39] We refrain from displaying and discussing all the details of his approach. Technically, his system is multi–**IKD**, where the (ten) operators are linearly ordered with respect to relative strength by axioms of the form $\Box_k\varphi \rightarrow \Box_\ell\varphi$ if $k > \ell$. $\Box_{10}$ corresponds to our A_i, and $\Box_5$ to our B_i operator.

- The scope of A_i in (53) is wrong; this is not due to a printing error or something like that, since wide scoping of A_i would produce a consistent formula, modulo the given system $\mathbf{IKD_{(10)}}$ and the ordering axioms. But the small scope of A_i in (53) indicates that only the first part of (23) is actually asserted when the sentence is uttered and the second part is not. It is hard to accept this judgement.
- A second problem for Åqvist's proposal arises in connection with sentences of the form (30). Since 'I assert that p' and 'I know that p' are logically independent, we do not obtain the desired contradiction. Alternatively, if we assume K_i to imply A_i (cf. our modality B_iK_i), we still cannot derive a contradiction within Åqvist's analysis: here the small scope of A_i causes the consistency of $A_ip \wedge \neg K_ip \wedge \neg K_i\neg p$.
- We find it hard to consider 'assertion' a sense of 'belief'. Of course not just the mental state, but also the act of utterance is involved in making an assertion. Strictly speaking both concepts are logically independent: only given some condition on the context of utterance, an assertion can involve some kind of belief, although presumably not the kind of (strict) belief Åqvist had in mind.[40]

Lenzen

Some time after completing a preliminary Dutch version of the paper underlying the present chapter,[41] Johan van Benthem brought [Le80, chapter 5] to our notice, which we had completely overlooked. Wolfgang Lenzen gives a solution to Moore's problem that is very similar to ours. With him, uttering a sentence φ involves being convinced that φ.[42] Now, 'being convinced' is analyzed as 'believing to know':

$$\ddot{U}_x\varphi \leftrightarrow B_xK_x\varphi.$$

The intimate connection between conviction (German: Überzeugung), and assertion is reflected in the following quote:[43]

> In der Tat scheint mir die [...] Äquivalenz zwischen Überzeugt–sein und Zu–wissen–glauben ebenso wie die mehrfach erwähnte pragmatische Bedingung der Bereitschaft zu einer Wissens–Behauptung *das* Charakteristikum des 'normalen', umgangssprachlichen Wissenbegriffs zu sein.

In all, his solution amounts to our UTT. His argumentation is partly different, though. On the one hand, Lenzen does not relate UTT to the Gricean maxims.[44] On the other

[40] A superior account of the relevance of various belief readings (involving different degrees of speaker's commitment) for the problems at stake is [Fu71, pp.231–247]. Furberg's semi-formal counterpart of (53) on [Fu71, p.247] has to be rejected for roughly the same reasons, though.

[41] 'Pragmatiek en doxastisch/epistemische logica', tot version, Tilburg University, Dpt. of Language, 1985

[42] [Le80, pp.130,131]

[43] [Le80, p.73]

[44] Lenzen does discuss a rule similar to our DOX and claims it to be derivable from the Gricean maxims, [Le80, p.129]. As we have seen, such a derivation is unlikely, since the most obvious (belief) maxim is too weak.

hand, Lenzen provides independent motivation by analyzing other types of epistemic paradoxes (such as the scepticist's, and the surprise examination) by means of, essentially, UTT. In conclusion, we sympathize with his approach in this matter.[45] Unlike Lenzen, however, we do not believe the operator $\ddot{U}$ to be more than a convenience. There are in fact several reasons for keeping B and K as the basic operators:

- $\ddot{U}$ can be defined in terms of B and K;
- B cannot be defined properly by $\ddot{U}$ and K;
- K cannot be defined properly by $\ddot{U}$ and B;
- the system based on $\ddot{U}$ and K is in danger of collapse.

The first point is obvious. The second point is rather that there seems to be no way to define the much weaker belief operator in terms of the stronger notions of conviction and knowledge. For example, taking duals of K or $\ddot{U}$, or taking belief to be $B\varphi = \ddot{U}\varphi \wedge \neg K\varphi$ would not lead to a normal modal logic for B. The third point is even more rhetorical: of course we can define knowledge as true conviction, as some authors are willing to do, but this does not imply that knowledge is based on sufficient evidence. The final point refers to the observation that systems based on K and $\ddot{U}$ are less stable than those based on K and B: we cannot simply require stronger properties of knowledge and belief without equating the two notions. For example, in [KL88] it is noticed that if the modal system of K_a is **S5**, and that of $\ddot{U}_a$ **NKD4**, $K_a\varphi$ implies $\ddot{U}_a\varphi$, and the typical axiom

Ü $\ddot{U}_a\varphi \to \ddot{U}_a K_a\varphi$

holds, then the logic collapses in the sense that $\vdash \ddot{U}_a\varphi \leftrightarrow K_a\varphi$.[46] We therefore conclude that a logic based on K and B is more economical, sufficiently rich, and easier to handle.

5.6 Extending the proposal

So far, we have only considered the effect of an utterance on the speaker himself. Although the rule UTT seems adequate, at least for the set of data discussed, it cannot be the full story. For the purpose of an utterance is usually not to make a self-belief explicit, but to try to convey a piece of information or conviction. So, utterances are usually part of a conversation, which implies the existence of a hearer different from the speaker. Now what is the effect of an utterance on the knowledge state of a hearer?

[45] This is not to imply that we fully sustain the logical systems developed in [Le80], which seem, intuitively spreaking, often too strong. For example, $\ddot{U}_x\varphi$ is supposed to be equivalent to $\hat{K}_x K_x\varphi$, which seems wrong to me. This may be related to the interpretation of $\widehat{\ddot{U}}_x$ as 'X considers it possible that' ('X hält es für möglich daß'), instead of the more accurate 'with respect to X's conviction it is possible that'.

[46] This collapse even occurs when the modal systems for K_a and $\ddot{U}_a$ are **NK5** and **NKD**, respectively. See e.g. [vdH91].

A first aspect of this *epistemic transfer* is that the hearer is convinced that the speaker is convinced of what he says, i.e.

$$i : `\varphi' \rightarrow B_j K_j B_i K_i \varphi$$

In other words, the rule UTT is available to the hearer, and thus 'internalized' by him, so to speak. Notice that several obvious alternatives, though more simple, do not qualify. For example, the formula $B_j B_i \varphi$ replacing the right-hand side of the rule of generalized conviction is too weak, since the hearer is usually reasonably sure of what he perceives. $B_j K_i \varphi$, by contrast, would be too strong in some sense, since it implies $B_j \varphi$. The proposed revision of UTT accounts for the fact that the Moore sentences are also strange to *hear*. In fact, the speaker in his turn may be convinced that this rule affects the hearer's mental state in the way described, i.e. $B_i K_i B_j K_j B_i K_i \varphi$. There may be no limit to the depth of mutual conviction: the next step would involve $B_j K_j B_i K_i B_j K_j B_i K_i \varphi$, and so on. Perhaps the conjunction of this sequence, suggesting a new notion of *common conviction*, comes to mind, but in general we do not see the need of such complex convictions in everyday conversations.[47] Moreover, this is only one side of the coin: all these convictions can eventually be traced down to the knowledge state of the speaker.

A second aspect, therefore, concerns the more direct effect of the information conveyed on the hearer: will he also be convinced that φ? That is simply asking too much — the hearer may hold opposite views. Also, it seems too much to require the speaker to merely believe that the transfer of *information* will always succeed. What seems usually defensible, though, is that it is logically possible, as far as the speaker believes, for the hearer to be convinced by what is said, i.e.

UTT2 $\quad i : `\varphi' \rightarrow \hat{B}_i B_j K_j \varphi.$ $\neg B_i \neg B_j K_j \varphi$

Notice that this rule refers to the situation occurring just after φ has been uttered. Before i utters φ it may very well be the case that j is not convinced that φ holds, and that i rightly believes so: $B_i \neg B_j K_j \varphi$. As a matter of fact this may be a very good reason for i to utter φ! Also, it may not even be likely that j is convinced by i with respect to φ; what is important is that this possibility is open in principle.

The second utterance rule UTT2 accounts for the anomaly of some second person variants of Moore's paradox, for example,

(54) p, but you do not believe that p.

(55) p, but you do not know whether p.

As with previous derivations, we can infer inconsistencies from the 'second order' pragmatic effects of rule UTT2 applied to the semantic representations of (54) and (55):[48]

---

[47] For perfect communication, the sequence of ever stronger utterance effects $K_j B_i K_i \varphi, K_i K_j B_i K_i \varphi, K_j K_i K_j B_i K_i \varphi, \ldots$ and its limit $CB_i K_i \varphi$ may be suggested. Again this knowledge seems too 'deep', but also too certain in colloquial language.

[48] Both paradoxes and UTT2 can be generalized to arbitrary subjects x overhearing the utterance. Simply replace j by such x.

(56) $\hat{B}_i B_j K_j(p \wedge \neg B_j p)$,

(57) $\hat{B}_i B_j K_j(p \wedge \neg K_j p \wedge \neg K_j \neg p)$.

Presumably (54) and (55) are less odd than their first person counterparts because 'believe' and 'know' may also refer to the situation (shortly) before the uttering takes place.[49] This is consistent with the *a posteriori* status of rule UTT2.

Again an alternative explanation is possible, which comes closer to *planning* an utterance: then the speaker is supposed to want to convince the hearer that φ. Notice this involves a partly *a priori* interpretation of the anomaly: one wants to convince before or while saying φ (the belief or knowledge still refers to the state after uttering). This would produce the rule

(58) $i : `\varphi' \rightarrow W_i B_j K_j \varphi$.

Recalling our representation of Grice's procedure P1 on page 137 we could favour another reformulation than (52) at this very point:

(59) $i : `\varphi' \rightarrow W_i B_j K_j B_i K_i \varphi$.

The latter two rules can jointly replace the combination of UTT and UTT2:

(60) $i : `\varphi' \rightarrow W_i B_j K_j(\varphi \wedge B_i K_i \varphi)$.

So far, we have found no empirical advantages for either the boulomaic or the purely doxastic/epistemic approach. For the time being we prefer the doxastic/epistemic line, since it does not invoke the additional and rather complicated operation of wanting.

Another issue not dealt with is the procedure for update of information, or, rather, update of the mental state. The above provides some rules for *adding* beliefs (or knowledge), but does not deal with the removal of contradictory beliefs. We consider this important but complicated issue to fall outside the scope of the present chapter.

5.7 Reinspecting the modal systems of B, K and $\ddot{U}$

The modal system behind the epistemic and doxastic logic presupposed in this chapter was that of [Hi62], i.e. essentially the multi-modal system in which K_a is **S4**, B_a is **NKD4**, and K_a implies B_a.

Before we turn to a reconsideration of the modal properties of B_a and K_a, let us investigate the logic of the modality BK, i.e. $\ddot{U}$. Assuming the above multi-modal system for B_a and K_a, what are the modal properties of $\ddot{U}_a$? It is easily verified that $\ddot{U}_a$ is at least **NKD4**, but is it more than that? Of course, if the operator K_a is still

[49] Interestingly, uttering (54) may sometimes be acceptable, whereas (55) is not. A possible explanation for this is that belief may be *irrational*, but knowledge is necessarily rational. Also, you can rationally believe someone else to be irrational, but you cannot rationally believe yourself to be irrational at that very moment.

available, there is a difference since Ü holds for $\ddot{U}_a$ but not for B_a (with respect to K_a). Without K_a, the logic of $\ddot{U}_a$ does not exceed **NKD4**: the class of Kripke models such that the accessibility relation for $\ddot{U}_a$ is the composition $\mathsf{B} \bullet \mathsf{K}$, where B is serial and transitive, K is reflexive and transitive, and $\mathsf{B} \subseteq \mathsf{K}$ is provably sound and complete for the system **NKD4**.[50] This shows the doxastic nature of $\ddot{U}_a$.

Although largely defended and taken over by other authors, Hintikka's modal system is not beyond criticism. In particular it can be criticized along four dimensions:

- the dimension of *logical omniscience*
- the dimension of *truth and consistency*
- the dimension of *introspection*
- the dimension of *interaction*

In section 5.3 we noticed that a great deal of omniscience can be removed by avoiding rule **N**.[51] But can we dispense with **N**, i.e. can we still derive the (pragmatic) inconsistencies within a logical system that lacks **N**? Fortunately we can, if **N** is replaced by the weaker rule **I**. Although **N** is presumably the gravest form of omniscience, **I**, **K** and, sometimes, **C!**, are also blamed for this. A more radical move is to pass to a form of partial modal logic which has no valid formulas at all. Again inspection of the relevant arguments shows that inference rules (instead of axioms) will do just as well, and that, moreover, the rule of *tertium non datur* ($\varphi \Rightarrow \psi \vee \neg\psi$) is not needed to derive the contradictions. Hence the given explanation is conserved when passing to this weaker logical system.

The truth axiom **T** for K_x was discussed at some length in section 5.2. It was supported by the semantic anomaly of (45) on page 135. The consistency axiom **D** is crucial for the explanation of Moore's paradox. Yet Lemmon, in his [Le65] review of [Hi62], noticed one can surely know or believe somebody else to have inconsistent beliefs: $K_a(B_b p \wedge B_b \neg p)$ should be consistent. This calls for a distinction of different sorts of belief: *rational* belief should be consistent, *possibly irrational* belief may be inconsistent. Consequently, if our intuitions concerning Moore's paradox are correct, it is, perhaps surprisingly, *rational belief* that is the more usual notion.

Whether we want to give up (positive) introspection depends on the sense of knowledge and belief we want to model. Idealized, implicit knowledge and belief may be considered to satisfy positive introspection. Intuitively, K and B are not quite the same in this respect. To us, it seems more sound to infer 'believing that you believe p' from 'believing that p', than to infer 'knowing that you know' from 'knowing' *simpliciter*, since 'believing that you believe' seems, in some sense, weaker

[50] Soundness is straightforward, and completeness follows from the fact that every non-**NKD4**-theorem has a counterexample which is a Kripke model with a serial and transitive accessibility relation R. This model can be transformed into one of the suitable form by taking $\mathsf{B} = R$ and $\mathsf{K} = R \cup id$, the reflexive closure of R. Then K is reflexive and transitive, and $\mathsf{B} \bullet \mathsf{K} = R \bullet (R \cup id) = R \bullet R \cup R \bullet id = R \bullet R \cup R = R$, since, by transitivity, $R \bullet R \subseteq R$.

[51] See chapters 6 and 7 for an extensive treatment of this issue.

than merely *believe*, whereas with knowledge it is the other way round. But then again, how can we know: the only clear intuitions seem to stem from those concepts we are reasonably aware of.

So possibly the **4** property for K and B should be avoided after all. Can we then still keep the analysis given in this chapter? Notice that we used **4** in deriving an inconsistency from $B_i(p \wedge \neg B_i p)$, to show that rule DOX was capable of handling the anomaly in uttering (22). Interestingly, rule UTT does not need **4** to produce an inconsistency from (22).

The other single place where **4** came in, was in establishing the relation between UTT and Grice's theory of conversational implicatures, more in particular, the maxim of quality.[52] Let there be no doubt that, if we were forced to give up either UTT or *quality*, we would prefer to keep the former. Alternatively, we might try to 'stretch' the maxim of belief further in such a way that uttering φ would involve $B_i\varphi \wedge B_iB_i\varphi$, but we do not think this is supported by intuitions or empirical arguments.

Avoiding **4**, there is at least one phenomenon we have to explain, viz. the anomaly in statements such as

(61) I believe that p, but I do not believe that I believe it.

(62) I know that p, but I do not know that I know it.

At first sight, the oddity of these examples provides independent evidence for **4**. Fortunately, UTT also explains why one cannot properly use these sentences: since they are of the form (22) and (46), a contradiction can be derived without using **4**.[53]

It may seem at this point that **4** can be avoided entirely. But things are not as simple as that: after all, the following examples seem to require **4** to logically account for the pragmatic inconsistency.

(63) p, but I do not believe that I believe it.

(64) p, but I do not know whether I know it.

This looks pretty much like an indirect argument for **4**. Another test case for the rule UTT is the logical order of the utterances of:

(65) I know it is raining.

(66) It is raining.

(67) I believe it is raining.

[52] As was noticed before, EPI does not have to be stronger than UTT, so there is no need to invoke **4** to establish such a consequence.

[53] For example, uttering (61) implies a contradiction by an argument that only uses (relativized) **ICXT**: $B_iK_i(B_ip \wedge \neg B_iB_ip) \Rightarrow B_iK_iB_ip \wedge B_iK_i\neg B_iB_ip \Rightarrow B_iB_iB_ip \wedge B_i\neg B_iB_ip \Rightarrow B_iB_iB_ip \wedge \neg B_iB_iB_ip$.

If UTT is correct and **4** holds, then the assertions of (65) and (66) should be equivalent, modulo the other logical rules. This does not appear to be right, so either **4** should be rejected, or another explanation (based on, for example, the Gricean maxims of *quantity* and *manner*) of the difference should be given. By contrast, asserting (66) seems to imply (67), and here positive introspection of knowledge is called for. However, perhaps we are relating the *semantics* of (67) to the *pragmatics* of (66). On the same level, and for weak, non-introspective senses of knowledge and belief, the latter two examples may indeed be independent.

With respect to the interaction of knowledge and belief, the axiom dubbed **X** here is widely accepted. Exceptions to this axiom involve admittedly irrational senses of belief such as in 'I know he is dead, yet I cannot believe it'. There is little need to strengthen the connection between knowledge and belief beyond **X**. We already provided some arguments *contra* scheme **Ü** for knowledge and plain belief (i.e. $B_x\varphi \rightarrow B_xK_x\varphi$). A similar scheme $B_x\varphi \rightarrow K_xB_x\varphi$ does not seem acceptable either.[54] We did use a strengthening of the **X** axiom to derive UTT from the Gricean maxim of *quality*. The identification of knowledge as true justified belief has been disputed by two diametrically opposed arguments: one point of criticism is that *justification* is not required for knowledge (cf. Gazdar's argument on page 138), another that even true justified belief may not suffice for knowledge. In Gettier's paradox[55] an agent believes a true proposition for which he has personal justification, but this evidence happens to be wrong. This allegedly shows that he has true justified belief, without really knowing it, and so (37) would be falsified. Yet to me, what the example shows is that the agent merely *believes* to have sufficient evidence, i.e. the situation seems correctly described by $Bp \wedge p \wedge BEp \wedge \neg Ep$, which is consistent with (37); applying (37) gives $p \wedge BKp \wedge \neg Kp$, modulo positive introspection.

Another issue concerning logical strength is also related to UTT. It was noticed that UTT solves Moore's paradox and its epistemic variants, yet does not seem too strong, logically speaking. This may be the proper place to notice that UTT is not the only, nay, not even the optimum solution with regard to logical strength: for example, the modality $\hat{K}K$ is superior in that it still accounts for Moore's examples, but is clearly weaker in the sense that

$$B_xK_x\varphi \Rightarrow \hat{B}_xK_x\varphi \Rightarrow \hat{K}_xK_x\varphi.$$

Moreover, $\hat{K}K$ has the obvious advantage that it is expressible within the uniform epistemic logic of K_x. Perhaps the modality $\hat{K}_xK_x$ was intended as one of the optional 'expectations' displayed in the quotation from [Hi62] on page 140. However, the weakest possible explanation may not be the best here. In fact, the choice of the logically stronger modality in UTT is supported by intuition. Also, the operators B_x, so there is hardly a real advantage of staying within the K_x system. Without feeling the need to discuss all possible modal solutions for the problem of the assertive force of indicatives (within certain bounds), we notice that some rather strong modalities,

[54] $B_x\varphi \rightarrow K_xB_x\varphi$ is accepted by [KL88], but rejected by Hintikka, see [Hi62, p.52].
[55] See [Le80]

such as $K_x B_x$ are not suitable: the latter one does not deal with the epistemic variants of Moore's paradox.

5.8 Conclusion

We have described and explained a number of challenging pragmatic anomalies connected to Moore's paradox. The rule which explains these phenomena is claimed to have a very general application to everyday statements. This utterance rule and the Gricean rule of *quality* are intimately connected. We have also argued that our rule is superior to many other solutions to Moore's puzzle. Our result is equivalent to a rule suggested (informally) by Lenzen, who arrives at his result by a somewhat different route (but from the same starting point, viz. Hintikka's work and Moore's paradox). The formalization given may be implemented in a compositional treatment of the pragmatic meaning of assertions, or rather their effect on knowledge and belief. Needless to say, this only constitutes an approximation of (one aspect of) pragmatic meaning.

Yet we argued that our utterance rule is adequate, both on the logical dimension and on the empiral dimension. It is logically correct, since it is not too strong and though itself consistent, can account for the observed pragmatic inconsistencies. It seems empirically correct since it fits the intuition and accords to Grice's maxim of quality, when the latter is modified in a way that seems necessary anyway.

The analysis is extended to second person variants of Moore's paradox. The final section indicates that the given explanation does not depend on a normal modal system; in other words, the explanation can be carried over to less idealized analyses of knowledge and belief.

Chapter 6

Total logics of awareness

This chapter[1] provides a study of total (bivalently interpreted) logics of belief and awareness, from a general modal perspective.

One form of possible world semantics proposed by Fagin & Halpern in [FH88] is generalized to what we call 'sieve semantics', which is essentially Kripke semantics with a superimposed awareness sieve. For each world, the awareness sieve specifies a set of formulas. In general these sets are arbitrary, but for special kinds of awareness we may constrain them by imposing conditions on the semantics. Sieve semantics turns out to provide an extremely flexible framework, which is proved to be effectively equivalent to Rantala's non-normal world semantics. Consequently, every awareness logic which contains the propositional tautologies can be given a sieve semantics.

After introducing some specific awareness logics, we study their completeness and correspondence properties. Then these logics are compared to each other, demonstrating their differences and similarities.

The positive aim of this part of the enterprise is an adequate description of awareness and actual belief, as well as the general model theory required, a 'negative' aim the avoidance of logical omniscience. These goals reflect two sides of the same coin, of course.

First we treat some modal logics in which awareness is essentially a syntactic filter on potential beliefs. Monotonicity conditions on the awareness filter then account for some kinds of active belief.

Next we treat neighbourhood semantics and related formalisms for active belief. In a topological metaphor, the different frames of mind which can be attributed to an agent may be construed as neighbourhood bases interpreting actual beliefs. Some alternatives to neighbourhood semantics are shown to be (almost) equivalent, yet a standard account is argued to be preferable.

[1]The present chapter is a modified version of [Th91a], ©Kluwer Academic Publishers 1992, with some deletions and extensions, such as a short discussion of [GG90].

6.1 Introduction

The ultimate goal of awareness logics is to give a sound and complete, yet descriptively adequate (modal) logic for awareness in its relation to human belief. Apart from academic interest this is also important for future communication systems which will be required to act as if they understand how humans think. Like [Pe90] we do not really expect machines to become conscious; however, by describing and implementing formal properties of consciousness we can make computers seem more intelligent and user-friendly.[2]

The present chapter is a step towards this goal. Although we will not give an empirically complete description of awareness and actual belief, a first approximation is provided. More importantly, a general framework is established which enables us to deal with various types of awareness and actual belief without the need to change the logic over and over again. Fortunately, since the framework covers earlier proposals (in fact generalizes one of them), previous insights are preserved and extended with more details, especially on (weak) introspection properties of active belief.

The impetus to what we call *awareness logics* are the problems of so-called 'logical omniscience'. This ironic term refers to the fact that standard logics such as (the minimal) *normal* modal logic fall short when they are applied to certain cognitive modes of human beings (or their simulations in AI). The problem is that these logics would force the agent to know or believe simply too much. More precisely, they would oblige a person to know all the consequences of his knowledge. For example, all number theorists now would 'know' whether Fermat's last theorem holds or not, since they know the postulates for ordinary arithmetic.[3] This is surely not the case, in any realistic sense of the word 'know': though these mathematicians may be said to *implicitly* know the answer to this classical query, nobody is aware of the answer, i.e. nobody knows it *explicitly*, so far. Or, more simply and perhaps even more convincingly, if somebody believes[4] p, he need not (explicitly) believe p *or* q, although any logic containing the classical tautologies and the principle

$$\text{I} \qquad \vdash \varphi \to \psi \Rightarrow \vdash B\varphi \to B\psi,$$

which modalizes the parts of a valid implication, would predict so.

To judge whether these problems were rightly ignored in [Hi62], we have to distinguish between *implicit* and *explicit belief*. I believe (explicitly, as a matter of fact) that Hintikka was trying to model implicit rather than explicit belief and was, therefore, virtually correct on this point.

Now it may seem easy to circumvent problems of logical omniscience (LO) by limiting the inferential power. Although this is precisely what awareness logics do, there are a fairly large number of complications to be dealt with. One is that there are many sorts of awareness and logical omniscience, and it appears to be difficult to

[2] See chapters 8 and 9.

[3] Assuming that the conjecture is not independent of Peano's axioms.

[4] In the sequel we will restrict ourselves to *belief*. Much of what will be said about belief also goes for *knowledge*, however.

deal with all of them at once. Apart from the forementioned principle **I**, some other prominent types of LO are:

N $\quad \vdash \varphi \Rightarrow \vdash B\varphi$

K $\quad \vdash B(\varphi \rightarrow \psi) \rightarrow (B\varphi \rightarrow B\psi)$

C $\quad \vdash (B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$

E $\quad \vdash \varphi \leftrightarrow \psi \Rightarrow \vdash B\varphi \leftrightarrow B\psi$

In an *Animal Farm*-like paradox, one may say that all types of LO are equally troublesome, but some are more troublesome than others, since these principles are ordered by the consequence series **NK** $\Rightarrow$ **I** $\Rightarrow$ **E** and **IK** $\Rightarrow$ **C**.

Another complication for building such logics is that of keeping classical propositional logic (**pL**) in the *external* part of the logic while avoiding omniscience in the *internal* part. For example, $Bp \vee \neg Bp$ should be valid but $B(p \vee \neg p)$ should not. For it seems obvious that anyone believes some fact or other, or he does not, and, most importantly, this holds regardless of the sense of belief involved. But it is not obvious at all that everyone should have any belief with regard to p at all, and therefore he need not explicitly believe the tautology *p or not p* either. Of course the axioms can be chosen in such a way that the modal system has the required effect; a minimal solution to this problem would involve just (the modal instantiations of) **pL**. Some weak principles such as the converse of **C**

$\mathbf{C_c} \quad \vdash B(\varphi \wedge \psi) \rightarrow (B\varphi \wedge B\psi)$

also seem fully acceptable for real belief. A complete specification, however, presupposes a clear-cut choice for the notion of *belief*, which unfortunately seems to have many appearances.

Now a chief difficulty is that removing LO–inference rules reopens the search for a suitable model theory, unless we want to abandon semantics altogether. Here some subtlety is required. For example, a straightforward partial logic which eliminates LO will also destroy **pL**.

It has even been suggested that possible worlds alone are to blame for LO. This sweeping statement is in fact not correct. As we have seen in chapter 4, partial semantics can equally well provide LO, since they may be used to describe normal modal logics such as **K** and **S4**,[5] depending on the manner of validation. Moreover, by adding certain modifications and generalizations, different types of LO may be avoided in possible world semantics, as this chapter will try to demonstrate.

The different approaches to solve problems of LO can be divided into:

- purely semantic approaches (i.e. containing a clean model theory, based on semantic intuitions) — this amounts to *weakening the logic*;

[5] I.e., under the perspective of *mixed falsifiability* on *coherent* models. Notice that, in some sense, [Hi62]'s model sets contain implicit partiality.

- strictly syntactic approaches (based on sets of formulas with limited deductive power), such as Konolige's;
- approaches with 'syntactically polluted' semantics, such as in the sieve models.

Since in our view a logic has to contain a proper semantics, we will disregard the second approach. Although the first approach usually involves an overtly *partial* semantics[6], to some extent it can be incorporated within the possible world paradigm; then it is related to neighbourhood semantics. In this chapter we mainly deal with the last approach. In its least radical forms one may say that partiality is simulated; in the more radical forms, we can remove virtually every modal principle, and thus eliminate every form of LO.

Apart from being based on some variant of possible world semantics, the theories that we will discuss have the common feature that explicit belief is connected to implicit belief by adding awareness to it. This can be put into a slogan:

EXPLICIT BELIEF = IMPLICIT BELIEF + AWARENESS

Like all slogans, the statement is rather imprecise. It does not account for the exact relationship, nor whether the notions are merely semantic or have a syntactic counterpart, nor which notion is primitive, and which one derived. In fact the latter may depend: sometimes explicit belief is derived, sometimes awareness is derived. Yet the equation points at a division of labour: the logical properties of actual belief may be thought to be present in idealized form in implicit belief, the non-logical character in the somewhat misty notion of awareness. Sometimes it is even claimed that awareness is an illogical notion. We feel that although it is true that awareness has no nontrivial properties in general (i.e., apart from missing ordinary properties), for special types we may and will formulate constraints.

6.2 Overview

The rest of the chapter is organized as follows. Sections 6.3, 6.4 and 6.5 deal with augmented Kripke semantics and section 6.6 with neighbourhood semantics.

We start with Fagin&Halpern's *general awareness logic* (GAL), and study its monotonicity behaviour. Konolige's criticism is discussed and some alternatives are compared to the original logic. Next the *special awareness logic* is introduced that describes a weak form of awareness related to familiarity with sufficiently many simple facts. Again monotonicity constraints turn out to be of vital importance for determining introspection properties of awareness and belief, although *negative introspection* with respect to explicit belief causes a collapse of the logic. We show that the latter logic may be considered as a special case of the general awareness logic, by embedding its semantics into that of the general one.

Then it is shown that a generalization of the GAL models, called sieve models, provides an extremely flexible semantic framework which is effectively equivalent

[6] See the next chapter.

to Rantala's non-normal world semantics. Sieve semantics thus covers all awareness logics which contain at least the classical propositional calculus.

Finally the general theory of neighbourhood models is given in order to deal with awareness within a frame of mind. It is argued that only very weak conditions should be imposed on these models. Two related, seemingly new forms of model theory are reduced to special kinds of neighbourhood semantics. Finally neighbourhood semantics and its variants are shown to be equivalent to branches of sieve semantics, though the latter is still more general than neighbourhood semantics.

Throughout the chapter a number of correspondence and completeness results will be mentioned or proved. In this way we can uniformize and improve some earlier proposals on this point.

6.3 The logic of general awareness

In order to obtain sufficient power to model awareness on the one hand and avoid problems of omniscience on the other, Fagin&Halpern (F&H henceforth) suggest the logic of general awareness. The non-logical nature of awareness is built into the logic by making the awareness set $\mathcal{A}_i(w)$ an arbitrary set of formulas. Roughly, awareness works like a sieve, filtering out explicit beliefs from the bulk of implicit beliefs. This is perhaps the most obvious realization of the equation given in the introduction.

The syntactic nature of awareness is reflected in the presence of a primitive awareness operator A_i for each agent i. To each of the m agents explicit (B_i) and implicit (L_i) beliefs are attributed.[7] So the language is essentially $\mathcal{L}_{\neg,\wedge,\{L_i\}_i,\{A_i\}_i,\{B_i\}_i}(Prop)$ (or $\mathcal{L}_{\vec{L},\vec{A},\vec{B}}$ for short), although B_i can also be introduced by the definition $B_i\varphi = L_i\varphi \wedge A_i\varphi$.

semantics

$\langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$ is a model of general awareness, if $\langle W, \vec{\mathsf{R}}, V\rangle$ is an ordinary (multi-modal) Kripke model in which W is a set of possible worlds[8], the accessibility relation $\mathsf{R}_i \subseteq W \times W$ (dealing with *implicit* beliefs) *serial*, *transitive* and *euclidian*, and V an ordinary two-valued valuation function, i.e. $V : Prop \times W \longrightarrow \{0,1\}$. Furthermore $\mathcal{A}_i(w) \subseteq \mathcal{L}_{\vec{L},\vec{A},\vec{B}}$ for all i, w, and the truth and validity conditions are standard-type apart from the non-recursive part caused by $\mathcal{A}$: [9]

- $M, w \models p$ iff $V(p, w) = 1$ ($p \in Prop$);
- $M, w \models \neg\varphi$ iff $M, w \not\models \varphi$;

[7]Throughout these sections we will sometimes use B instead of B_i, etcetera; within a formula or rule, modal operators are to be considered as coindexed by default.

[8]Or states, as F&H call them.

[9][FH88] have S and s where we have W and w, π for our V, B for R, Φ for *Prop*, **true** for 1, **false** for 0, ***true*** for $\top$, ***false*** for $\bot$, $\sim$ for $\neg$, $\Rightarrow$ for $\rightarrow$, and $\equiv$ for $\leftrightarrow$.

- $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$;
- $M, w \models L_i\varphi$ iff $M, v \models \varphi$ for every v such that $w\mathsf{R}_i v$;
- $M, w \models A_i\varphi$ iff $\varphi \in \mathcal{A}_i(w)$;
- $M, w \models B_i\varphi$ iff $\varphi \in \mathcal{A}_i(w)$ and $M, v \models \varphi$ for every v such that $w\mathsf{R}_i v$;
- $\models \varphi$ iff $M, w \models \varphi$ for all models M and worlds w in M.

completeness and correspondence

This simple semantics enables a nice and easy completeness result:

Theorem 6.1 (Fagin&Halpern) [10]
The modal system for the logic of general awareness is **pL + weak S5** *(i.e.* **NKD45***)*[11] *for L_i and the axiom $\vdash B_i\varphi \leftrightarrow L_i\varphi \wedge A_i\varphi$.*

Proof: Soundness is obvious and completeness is shown by a straightforward Henkin proof. Canonical awareness sets are defined by $\mathcal{A}_i(\Sigma) = A_i{}^{-1}[\Sigma]$.[12] ■

Despite its simplicity the framework is a very flexible tool: different types of awareness and explicit belief are easy to model, as demonstrated by a number of correspondences. Here *monotonicity* constraints enter our story.

Definition 6.1 (monotonicity)

(mon↑) $\mathcal{A}_i$ *is* **upwards monotone** *with respect to* R_i *iff* $w\mathsf{R}_i v \Rightarrow \mathcal{A}_i(w) \subseteq \mathcal{A}_i(v)$ *for all* w, v;

(mon↓) $\mathcal{A}_i$ *is* **downwards monotone** *with respect to* R_i *iff* $w\mathsf{R}_i v \Rightarrow \mathcal{A}_i(v) \subseteq \mathcal{A}_i(w)$ *for all* w, v;

(mon=) $\mathcal{A}_i$ *is* **constantly monotone** *with respect to* R_i *iff* $w\mathsf{R}_i v \Rightarrow \mathcal{A}_i(w) = \mathcal{A}_i(v)$ *for all* w, v.

1. Introspection (with respect to awareness) amounts to the axiom $A\varphi \rightarrow AA\varphi$. This corresponds to the condition $A[\mathcal{A}(w)] \subseteq \mathcal{A}(w)$ on structures of general awareness.
2. Upward monotonicity corresponds to the axiom $A\varphi \rightarrow LA\varphi$.
3. Downward monotonicity corresponds to the axiom $\neg A\varphi \rightarrow L\neg A\varphi$.

---

[10] [FH88, theorem 8.4]

[11] **D** stands for $\vdash L_i\varphi \rightarrow \neg L_i\neg\varphi$, **4** for $\vdash L_i\varphi \rightarrow L_iL_i\varphi$, and **5** for $\vdash \neg L_i\varphi \rightarrow L_i\neg L_i\varphi$.

[12] I.e. $\varphi \in \mathcal{A}_i(\Sigma) \Leftrightarrow A_i\varphi \in \Sigma$.

4. Closure of all $\mathcal{A}(w)$ under subformulas (within the 'small' language $\mathcal{L}_{\vec{L},\vec{A}}$) corresponds to the addition of the axioms:[13]

$$\vdash (AL\varphi \vee AA\varphi \vee A\neg\varphi) \rightarrow A\varphi,$$
$$\vdash A(\varphi \wedge \psi) \rightarrow (A\varphi \wedge A\psi).$$

5. Next consider the case of awareness of a set of atoms, or rather, of all formulas containing just those atoms. I.e., for all worlds and every agent there is a subset $\Psi \subseteq Prop$ such that $\mathcal{A}(w) = \mathcal{L}(\Psi)$. In the small language (with B_i introduced by definition), this corresponds to the system consisting of the axioms $\vdash AL\varphi \leftrightarrow A\varphi, \vdash AA\varphi \leftrightarrow A\varphi, \vdash A\neg\varphi \leftrightarrow A\varphi$ and $\vdash A(\varphi \wedge \psi) \leftrightarrow (A\varphi \wedge A\psi)$.

6. Finally consider awareness by limited time/space bounds, for example (local) knowledge of a processor (i) in a distributed system.[14] This boils down to requiring *reflexivity* of the relations R_i and constant monotonicity of $\mathcal{A}_i$ with respect to R_i. The corresponding axioms are $L_i\varphi \rightarrow \varphi$, $A_i\varphi \rightarrow L_iA_i\varphi$ and $\neg A_i\varphi \rightarrow L_i\neg A_i\varphi$.

monotonicity effects and evaluation

The generality of this approach enables instantiations for special kinds of awareness. Here are some comments on these special kinds.

First, with respect to monotonicity, we can inspect the monotone kinds of general awareness emphasizing the bilateral relationships among awareness, implicit and explicit belief.

Proposition 6.1 *The implications displayed below are valid under mon↑, their converses under mon↓, and their bidirectional counterparts (i.e. equivalences) under mon=:*

$$\begin{array}{ll}
\models_{mon\uparrow} A\varphi \rightarrow LA\varphi & \models_{mon\uparrow} L\neg A\varphi \rightarrow \neg A\varphi \\
\models_{mon\uparrow} B\varphi \rightarrow LB\varphi & \models_{mon\uparrow} L\neg B\varphi \rightarrow \neg B\varphi \\
\models_{mon\uparrow} (B\varphi \wedge AB\varphi) \rightarrow BB\varphi & \models_{mon\uparrow} B\neg B\varphi \rightarrow (\neg B\varphi \wedge A\neg B\varphi)
\end{array}$$

Proof: straightforward from the definitions. ∎

So, in the light of what follows later on, noteworthy instances are $\models_{mon\uparrow} B\varphi \rightarrow LB\varphi$, $\models_{mon\downarrow} LB\varphi \rightarrow B\varphi$, $\models_{mon=} LB\varphi \leftrightarrow B\varphi$, $\models_{mon\uparrow} L\neg B\varphi \rightarrow \neg B\varphi$, $\models_{mon\downarrow} \neg B\varphi \rightarrow L\neg B\varphi$, and $\models_{mon=} L\neg B\varphi \leftrightarrow \neg B\varphi$.

Second, closure under subformulas, and in particular for conjunctions is allegedly motivated by reference to the 'pragmatically paradoxical' [15] formula $B(p \wedge \neg Bp)$ which would become satisfiable without imposing the restriction. In fact, as pointed

[13] Likewise extended for the full language by axioms such as $\vdash AB\varphi \rightarrow A\varphi$.

[14] See [FH88, p.57] for details.

[15] Cf. chapter 5.

out in [vdHM88], something more is needed to make $\neg B(p \wedge \neg Bp)$ valid: *upward monotonicity* should hold as well.[16] The price to pay is that **K**- and **C**-omniscience are regained, which is sometimes considered problematic for explicit belief.

Third, for resource-bounded reasoning it may not be obvious that the ambitious claims of F&H can be effectuated. In fact only the case for distributed computations is dealt with in some detail. F&H also mention cryptography as one of the possible applications. The feasibility of this type of application depends on the precise nature of the source: is it simply the length of the formulas or some other notion of syntactic complexity, the size of the possible models, the number of steps of the derivation, or what? It is perfectly clear that structural limitations (complexity of formulas) will be easier to incorporate than derivational limitations. Yet recent work on so-called *zero-knowledge* proofs[17] attempts to capture resource-bounded reasoning for cryptographic applications.

Fourth, Konolige claims that the semantics proposed for general awareness is not adequate since "the formal correspondence between accessibility relations and sets of awareness sentences breaks down" [Ko86, p.246]. Now, literally this is not the purpose of the awareness sets. So, let us try to make sense out of this quote by transforming it into a question: 'are there axiom schemes that have no formal correspondent in terms of a structural constraint on validating frames?' In fact let us consider the **4** axiom for B: $B\varphi \rightarrow BB\varphi$. Konolige apparently suggests that this scheme is troublesome. It is not difficult to formulate validating conditions: $\mathcal{A}$ has to be *mon*$\uparrow$ and closed with respect to B (i.e. $\varphi \in \mathcal{A}(w) \Rightarrow B\varphi \in \mathcal{A}(w)$). To get full correspondence we have to relax these conditions a bit: *mon*$\uparrow$ and B-closure only have to hold for formulas in worlds where they are modally satisfiable.[18] Now such a condition is rather unusual, and perhaps this may have worried Konolige, but this is not a formal reason to abandon it. A similar story goes for the **5** axiom of negative introspection of explicit belief.

So most of these specialities seem quite robust, but there has been some criticism on the general part of the story as well. [Ko86] holds that the logic is essentially the syntactic approach[19] in disguise. We basically disagree. True, the logic contains a large syntactic component: the awareness sets consist of (uninterpreted) formulas. Awareness thus becomes a (generally non-recursive) non-logical notion. But there is also a recursive semantics attached to it, dealing with ordinary logical aspects. This may seem a rather eclectic approach, combining syntactic and semantic elements, but here it is precisely what we want: the limited inference is accounted for by a proper semantics. Konolige wants to abolish the semantics altogether, but this only makes the logic less insightful — if we would express the inferences by deductive rules only, we have failed to give a reason for the properness of the inference. By its axiomatizations, the logic of general awareness is able to provide the deductive rules as well. Of

[16] Cf. [FH88, footnote 6]. Instead of *mon*$\uparrow$, [vdHM88, p.29] require the stronger *mon*$=$.

[17] The key reference is [GMR85]; see also [HMT88].

[18] I.e. for all such frames F, if there is a V such that $F, V, w \models L\varphi$ and $\varphi \in \mathcal{A}(w)$, then both $B\varphi \in \mathcal{A}(w)$ and $\varphi \in \mathcal{A}(v)$ for all v such that $w\mathsf{R}v$.

[19] The syntactic approach, which abolishes all model theory, is strictly based on sets of formulas and (limited) inference rules.

course, a purely semantic and fully recursive approach would be preferable, but we believe this is intrinsically impossible, due to the psychological nature of awareness: consciousness of the whole need not imply consciousness of all parts, and *vice versa*.

At this point it may be wise to point at some unanswered questions about axiomatizations for sublanguages. It is not obvious which modal systems trigger the various sublanguages. In order to gain insight into the logical properties of explicit belief: what is B's own system? What is the complete logic using merely B and A? And what is the one for just B and L? So far we have only partial answers to these queries. For example, some valid principles not mentioning L are:

- $\vdash \bigvee \neg\varphi_i \Rightarrow \vdash \bigvee \neg B\varphi_i$ **(N*)**
- $\vdash (B(\varphi \to \psi) \wedge B\varphi \wedge A\psi) \to B\psi$ [20]
- $\vdash B\varphi \to \neg BB\neg\varphi$
- $\vdash (B\varphi \wedge AB\varphi \wedge BA\varphi) \to BB\varphi$
- $\vdash (A\neg B\varphi \wedge B\neg A\varphi) \to B\neg B\varphi$

Notice that many of these principles resemble the usual axioms and rules of normal modal logic with enough awareness built-in. Also notice that **N*** implies **D** ($\vdash B\varphi \to \neg B\neg\varphi$), as well as **D*** ($\vdash \neg B(\varphi \wedge \neg\varphi)$) and other dual relaxations of principles that are invalid on their own, such as $\vdash (B\varphi \wedge B\psi) \to \neg B\neg(\varphi \wedge \psi)$, the weakened counterpart of **C**.

6.3.1 two specialized alternatives

Huang & Kwast

In [HK91] a special variant of the logic of general awareness is proposed. The characteristic features of this system are:

- certain conditions on the awareness functions $\mathcal{A}$:
  - *propositional closure*: $\neg\varphi \in \mathcal{A}(w) \Leftrightarrow \varphi \in \mathcal{A}(w)$ and $(\varphi \wedge \psi) \in \mathcal{A}(w) \Leftrightarrow \varphi, \psi \in \mathcal{A}(w)$,
  - *nested awareness*: $A\varphi \in \mathcal{A}(w) \Rightarrow \varphi \in \mathcal{A}(w)$,[21]
  - *belief awareness interpretation*: $A\varphi \in \mathcal{A}(w) \Leftrightarrow L\varphi \in \mathcal{A}(w)$;
- a modified definition of explicit belief (B_i):

$$B_i\varphi = L_i\varphi \wedge A_iL_i\varphi.$$

[20] This property was suggested in the chapter of Halpern et. al.'s forthcoming book, cf. note 36.

[21] The name of this condition stems from the axiom it triggers: $AA\varphi \to A\varphi$.

Some comments may be in order. To start with the last point, notice that, even without A_i, the new B_i (say B_i^{HK}) can be expressed in terms of in the old B_i (B_i^{FH}) and L_i, but not the other way round:

$$\models B_i^{\mathrm{HK}}\varphi \leftrightarrow B_i^{\mathrm{FH}} L_i\varphi.$$

So F&H's definition is more general. Perhaps the new definition is more adequate, but this does not follow from the alleged argument:

> However, according to the [F&H] definition, it is possible that [...] agent i may not be aware of φ in a[n accessible] state t though he believes φ is true. [HK91, p.294]

With regard to the above definition the quoted argument is a *non sequitur*: instead of implying $B_i\varphi \to A_i L_i\varphi$, it establishes $B_i\varphi \to L_i A_i\varphi$, which is validated by *mon*↑ of $\mathcal{A}_i$ (proposition 6.1).

With regard to the constraints on the awareness function $\mathcal{A}_i$, I am willing to believe the conditions of *propositional closure* and *nested awareness*[22], but it is difficult to accept the *belief awareness interpretation*. It is especially difficult to grasp the intuition behind $A_i L_i\varphi \Rightarrow A_i A_i\varphi$.

One of the prime goals of [HK91] is to show that by means of the awareness operator one can define different kinds of implications, some of which would lead to **K**- and **I**-type omniscience and others not. Yet only part of this aim has been realized.[23]

Gillet & Gochet

One complaint which is sometimes directed against GAL is that the awareness sets seem to be unstructured wholes, whereas at least one way of knowing and becoming aware of things is by thinking about them, by 'mental computation'. However, a more structured notion of awareness can easily be incorporated into GAL. One proposal, capturing *resource bounded* awareness was executed in [Mo88]. Another interesting suggestion was made in [GG90], where the levels of awareness are not based on computational limitations, but on the syntactic complexity of the formula. Some essential features of this system are:

- the depth d (French: 'profondeur') of a formula is, roughly, the maximum number of nested logical constants (including the belief operators) in a formula, with the exception of $\neg$ (which does not count) and $\leftrightarrow$ (which counts twice).

[22] Although we would still prefer the constraint of closure under subformulas, joined with the monotonicity conditions.

[23] Some of the reported results are not entirely correct. For example, proposition 6.2(c), which is important because it serves as a redefinition for the *strong implication* $\leadsto_i$, should read: $(\varphi \leadsto_i \psi) \leftrightarrow ((\varphi \to \psi) \wedge A_i\varphi \wedge A_i\psi \wedge A_i A_i \neg\psi)$ [the negation in the last conjunct was left out; the second last conjunct is not necessary for the equivalence but is vital for the new definition: it ensures **K**-closure]. Propositions 6.7 (b) and (c), which are presumably intended to express **I**-omniscience, are misleading: they hold vacuously since the premises are false.

- the awareness sets do not depend on the world under consideration, but on the depth of the formulas. In fact [GG90] do not use a more or less 'semantic' notion of awareness $\mathcal{A}^{(d)}$, but directly define awareness-up-to-depth d by means of a purely syntactic operation on formulas.[24] Instead of giving the formal definition, which is rather involved, we illustrate this part of the system by an example below.

- a definition of explicit belief-up-to-depth d, i.e. $B_i^{(d)}$:[25]

$$B_i^{(d)}\varphi = L_i A_i^{(d)}\varphi.$$

To illustrate the operation $A_i^{(d)}$, consider the following example. On the 0[th] level of awareness, a formula such as $p \to (p \vee q)$ is construed as being entirely opaque, i.e. as a new atom $a_{p\to(p\vee q)}$. On awareness level 1, the outermost structure of the formula is revealed: $a_p \to a_{p\vee q}$. On level 2 and higher, more and more information is revealed; here it reproduces the original formula $p \to (p \vee q)$.

The stratified notion of awareness licenses a similarly layered truth condition for $B^{(d)}$. Then valid depth d belief implies the validity of 'deeper' belief. Although the definitions of *depth* and *awareness* have some peculiar features (effecting, for example, the questionable validity of $B^{(d)}\neg\neg\varphi \leftrightarrow B^{(d)}\varphi$), the idea of layered awareness warrants further research.

6.3.2 the logic of special awareness

The *special awareness logic* (SAL) in [FH88] describes the type of explicit belief that can be related to a number of acquainted facts. In a sense, this may be regarded as a recursive alternative to GAL. In this set-up, there is no need for an awareness operator in the basic language, which can be characterized as $\mathcal{L}_{\vec{B},\vec{L}}$. A syntactic counterpart to the semantically present awareness can be defined, but it will lack the simple properties of A_i in GAL.

semantics

Models are of the form $\langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$, where, as in GAL, $\langle W, \mathsf{R}_i, V\rangle$ is an ordinary (**weak S5**) Kripke model which is augmented with awareness sets that are now sets of *propositional atoms*: $\mathcal{A}_i(w) \subseteq Prop$ for all i, w.[26]

Although the truth assignment to propositional atoms is classical, a *partial* effect is reached by restricting truth and falsity of formulas in worlds by means of the awareness sets. Here $\models^{\Psi}$ denotes truth with respect to Ψ, where Ψ is a set of conscious atoms. Likewise, $=\!\mid^{\Psi}$ stands for *falsity* with respect to Ψ. So $w \not\models^{\Psi} p$ does not imply

[24]Notice that this operator $A^{(d)}$ is *not* a modal operator of the logical language.

[25]At least, this is the effect of the truth condition for $B_i^{(d)}$; as far as we can see, the given redefinition accords to the other proposed definitions, and simplifies matters considerably.

[26]Again we have changed the notation. F&H's $\models_T^{\Psi}$ is replaced by $\models^{\Psi}$ and $\models_F^{\Psi}$ by $=\!\mid^{\Psi}$.

$w \dashv^{\Psi} p$, if $p \notin \Psi$. Apart from these restricted truth and falsity relations there is also an unrestricted classical truth relation ($\models$). These relations are defined by recursion:

- $M, w \models^{\Psi} p$ iff $V(p, w) = 1$ and $p \in \Psi$ (where $p \in Prop$);
  $M, w \dashv^{\Psi} p$ iff $V(p, w) = 0$ and $p \in \Psi$ ($p \in Prop$);
  $M, w \models p$ iff $V(p, w) = 1$ ($p \in Prop$);
- $M, w \models^{\Psi} \neg\varphi$ iff $M, w \dashv^{\Psi} \varphi$;
  $M, w \dashv^{\Psi} \neg\varphi$ iff $M, w \models^{\Psi} \varphi$;
  $M, w \models \neg\varphi$ iff $M, w \not\models \varphi$;
- $M, w \models^{\Psi} \varphi \wedge \psi$ iff $M, w \models^{\Psi} \varphi$ and $M, w \models^{\Psi} \psi$;
  $M, w \dashv^{\Psi} \varphi \wedge \psi$ iff $M, w \dashv^{\Psi} \varphi$ or $M, w \dashv^{\Psi} \psi$;
  $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$;
- $M, w \models^{\Psi} L_i\varphi$ iff $M, v \models^{\Psi} \varphi$ for every v such that $w\mathsf{R}_i v$;
  $M, w \dashv^{\Psi} L_i\varphi$ iff $M, v \dashv^{\Psi} \varphi$ for some v such that $w\mathsf{R}_i v$;
  $M, w \models L_i\varphi$ iff $M, v \models \varphi$ for every v such that $w\mathsf{R}_i v$;
- $M, w \models^{\Psi} B_i\varphi$ iff $M, v \models^{\Psi\cap\mathcal{A}_i(w)} \varphi$ for every v such that $w\mathsf{R}_i v$;
  $M, w \dashv^{\Psi} B_i\varphi$ iff $M, v \dashv^{\Psi\cap\mathcal{A}_i(w)} \varphi$ for some v such that $w\mathsf{R}_i v$;
  $M, w \models B_i\varphi$ iff $M, v \models^{\mathcal{A}_i(w)} \varphi$ for every v such that $w\mathsf{R}_i v$;
- $\models \varphi$ iff $M, w \models \varphi$ for all models M and world w in M.

Equipped with these definitions a kind of *persistence* can be shown.

Proposition 6.2 (Fagin&Halpern) [27]
If $\Psi \subseteq \Psi' \subseteq Prop$ then for all M, w, φ:

- $M, w \models^{\Psi} \varphi \Rightarrow M, w \models^{\Psi'} \varphi \Rightarrow M, w \models \varphi$
- $M, w \dashv^{\Psi} \varphi \Rightarrow M, w \dashv^{\Psi'} \varphi \Rightarrow M, w \not\models \varphi$

In particular, we have that $w \models^{\Psi} \varphi \Rightarrow w \models \varphi$, but not *vice versa*: the equivalence breaks down with the falsity case for B_i.[28] Proposition 6.2 is a convenient tool for proving validities.

[27] [FH88, proposition 4.1(2,3)]

[28] However, [Wa90, lemma 1] seemingly strengthens this result, that is, if we recursively specify the atoms in a formula φ relevant for awareness in w by means of the sets $\mathcal{T}_w(\varphi)$ and $\mathcal{F}_w(\varphi)$, and generalize his restriction set $\mathcal{A}_i(w')$ to an arbitrary $\Psi \subseteq Prop$ (to keep the induction going), there is an alleged converse:

$$M, w \models^{\Psi} \varphi \Leftrightarrow M, w \models \varphi \;\&\; \mathcal{T}_w(\varphi) \subseteq \Psi,$$

and likewise for $\dashv^{\Psi}$. Unfortunately, the claim is wrong: taking $\Psi = Prop$ and $\mathcal{A}_i(w) = \emptyset$ provides a singleton counter-example for $\varphi = \neg Bp$ (and similarly with $\mathcal{A}_i(w')$ instead of Ψ).

Proposition 6.3

$$\begin{array}{lll} \models BB\varphi \to B\varphi & \models B\varphi \to L\varphi & \models L\varphi \leftrightarrow LL\varphi \\ \models B\neg\varphi \to \neg B\varphi & \models B\varphi \leftrightarrow BL\varphi & \models L\neg\varphi \to \neg L\varphi \\ \models B\neg B\varphi \to \neg B\varphi & \models B\neg B\varphi \to \neg L\varphi & \models L\neg L\varphi \leftrightarrow \neg L\varphi \end{array}$$

Proof: R is not only transitive, but in addition serial and euclidian, and therefore dense. So, for example, $M, w \models BB\varphi \Rightarrow$ for each v and u such that $w\mathsf{R}v$ and $v\mathsf{R}u$: $M, u \models^{\mathcal{A}(w)\cap\mathcal{A}(v)} \varphi \Rightarrow$ for each u such that $w\mathsf{R}u$: $M, u \models^{\mathcal{A}(w)} \varphi \Rightarrow M, w \models B\varphi$. The other cases are similar. ■

Although awareness is not syntactically present, it can be reintroduced. The idea is that somebody may be said to be aware of (or acquainted with) a simple fact p, if p is true or false relative to the awareness set in every state he considers possible; in other words, if he explicitly believes $p \vee \neg p$. This is licensed by the observation that $p \in \mathcal{A}_i(w) \Rightarrow w \models B_i(p \vee \neg p)$. Likewise, one is aware of a complex fact if one is aware of all the primitives it contains. This suggests the definition

$$A\varphi = \bigwedge_{p \text{ in } \varphi} B(p \vee \neg p)$$

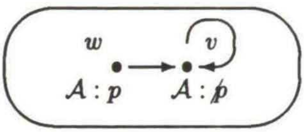
The following proposition relates explicit belief and awareness in the simple case of formulas that are free from modal operators.

Proposition 6.4 (Fagin&Halpern)
If φ is purely propositional and $\models \varphi$, then $\models A\varphi \to B\varphi$.

Proof: a simple contraposition argument suffices. ■

Notice that the restriction to *propositional* φ is essential. Without it proposition 6.4 does not hold any longer. Here is a counterexample:

Example 6.1 *Consider the model $\langle\{w, v\}, \mathsf{R}, \mathcal{A}, V\rangle$ where $\mathsf{R} = \{\langle w, v\rangle, \langle v, v\rangle\}$, $p \in \mathcal{A}(w)$, but $p \notin \mathcal{A}(v)$; V is arbitrary.*



Now let $\varphi = Bp \vee \neg Bp$. Then (i) $\models \varphi$. Furthermore, $A\varphi = B(p \vee \neg p)$, and $p \in \mathcal{A}_i(w) \Rightarrow v \models^{\mathcal{A}_i(w)} p \vee \neg p \Rightarrow w \models B(p \vee \neg p)$, hence (ii) $w \models A\varphi$. Finally, $p \notin \mathcal{A}(w)\cap\mathcal{A}(v)$ and $v\mathsf{R}v$, thus $v \not\models^{\mathcal{A}(w)} Bp$ and $v \not\dashv^{\mathcal{A}(w)} Bp$, so $v \not\models^{\mathcal{A}(w)} Bp\vee\neg Bp$, whence by $w\mathsf{R}v$, (iii) $w \not\models B\varphi$.

completeness

To obtain completeness [FH88] need a rather peculiar axiom

ANF $\quad \vdash \varphi \leftrightarrow \varphi^*$,

where φ^* is a normal form of φ in which each B_i can only have scope over $p \vee \neg p$ for atoms p occurring in φ. This amounts to $\varphi^* \in \mathcal{L}_{\vec{A},\vec{L}}$. Let us call such a φ^* an awareness normal form (ANF) of φ. F&H provide a procedure to derive an ANF for each φ in the original language. Though intricate, the procedure is entirely syntactic. Therefore it is possible to put thc equivalence of φ and its ANF into an axiom. Notice this axiom is extremely forceful and enables a succinct axiomatization.

Theorem 6.2 (Fagin&Halpern) [29]
The modal system for the logic of special awareness consists of **pL** *(including* MP*),* **weak S5** *(i.e.* **NKD45***) for L_i and the* ANF *axiom.*

Proof: Completeness is shown by a standard Henkin-style argument in which canonical worlds are maximally consistent sets and the canonical awareness functions are defined by $p \in \mathcal{A}_i(\Sigma) \Leftrightarrow B_i(p \vee \neg p) \in \Sigma$. Soundness is unusually difficult to prove, due to the ANF axiom. ■

monotonicity effects

Notice that the above counter-example can be eliminated by the fairly natural condition of *upward monotonicity* of awareness. After all, it seems rather plausible that once you are aware of p in some world, you still arc aware of p in each alternative you consider conceivable. The expectation that the generalization of proposition 6.4 is restored by this condition turns out to be right. However, with monotonicity we can prove a stronger result, directly relating the concepts of awareness, implicit and explicit belief. There is one proviso here: the formulas should contain operators related to one single agent (see example 6.2.)

Proposition 6.5 (one agent) $\models_{mon\uparrow} (A\varphi \wedge L\varphi) \rightarrow B\varphi$

This fact, which shows a partial similarity with GAL, is proved by means of a lemma.

Lemma 6.1 *For all monotone single-agent models $M = \langle W, \mathsf{R}, \mathcal{A}, V\rangle$, worlds w and v, and formulas φ such that $M, w \models A\varphi$ and $w\mathsf{R}v$:*

$$M, v \models^{\mathcal{A}(w)} \varphi \Leftrightarrow M, v \models \varphi \qquad M, v \dashv^{\mathcal{A}(w)} \varphi \Leftrightarrow M, v \not\models \varphi$$
$$M, v \models^{\Psi \cap \mathcal{A}(w)} \varphi \Leftrightarrow M, v \models^{\Psi} \varphi \qquad M, v \dashv^{\Psi \cap \mathcal{A}(w)} \varphi \Leftrightarrow M, v \dashv^{\Psi} \varphi$$

[29] [FH88, theorem 8.2], *vide* [l.c., pp.65,66] for an elaborated completeness proof and [l.c., pp.70-74] for soundness of the ANF axiom.

Proof: By a laborious simultaneous induction on the structure of φ. Let M be a model and w, v worlds such that $w\mathsf{R}v$ and $M, w \models A\varphi$ (we omit M in the rest of this proof). We show the key modal step where φ is of the form $B\psi$. Assume the lemma to hold for ψ (IH). $w \models AB\psi$, thus then by the definition of A: $w \models A\psi$ (i.e. for all p in ψ: $p \in \mathcal{A}(w)$). We need to prove four equivalences: (notice that the use of IH in the subsequent cases is triggered by transitivity of R)

- $v \models^{\mathcal{A}(w)} B\psi \Leftrightarrow \forall u \in \mathsf{R}[v] : u \models^{\mathcal{A}(w) \cap \mathcal{A}(v)} \psi \Leftrightarrow$ (IH) $\forall u \in \mathsf{R}[v] : u \models^{\mathcal{A}(v)} \psi \Leftrightarrow v \models B\psi$.
- $(\Rightarrow)$ $v \mathrel{=\!|}^{\mathcal{A}(w)} B\psi \Rightarrow \exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\mathcal{A}(w)\cap\mathcal{A}(v)} \psi \Rightarrow$ (IH) $\exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\mathcal{A}(v)} \psi \Rightarrow v \mathrel{=\!|} B\psi \Rightarrow v \not\models B\psi$. $(\Leftarrow)$ Suppose $v \not\models B\psi$, then $\exists u \in \mathsf{R}[v] : u \not\models^{\mathcal{A}(v)} \psi \Rightarrow$ (by mon$\uparrow$ + proposition 6.2) $\exists u \in \mathsf{R}[v] : u \not\models^{\mathcal{A}(w)} \psi \Rightarrow$ (IH) $\exists u \in \mathsf{R}[v] : u \not\models \psi \Rightarrow$ (IH) $\exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\mathcal{A}(w)} \psi \Rightarrow$ (mon$\uparrow$ + proposition 6.2) $\exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\mathcal{A}(w)\cap\mathcal{A}(v)} \psi \Rightarrow v \mathrel{=\!|}^{\mathcal{A}(w)} B\psi$.
- $v \models^{\Psi\cap\mathcal{A}(w)} B\psi \Leftrightarrow \forall u \in \mathsf{R}[v] : u \models^{\Psi\cap\mathcal{A}(w)\cap\mathcal{A}(v)} \psi \Leftrightarrow$ (IH) $\forall u \in \mathsf{R}[v] : u \models^{\Psi\cap\mathcal{A}(v)} \psi \Leftrightarrow v \models^{\Psi} B\psi$, and finally
- $v \mathrel{=\!|}^{\Psi\cap\mathcal{A}(w)} B\psi \Leftrightarrow \exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\Psi\cap\mathcal{A}(w)\cap\mathcal{A}(v)} \psi \Leftrightarrow$ (IH) $\exists u \in \mathsf{R}[v] : u \mathrel{=\!|}^{\Psi\cap\mathcal{A}(v)} \psi \Leftrightarrow v \mathrel{=\!|}^{\Psi} B\psi$. ■

Proof of proposition 6.5: immediately from lemma 6.1: if $M, w \models A\varphi$ and $M, w \models L\varphi$, then for any v such that $w\mathsf{R}v$: $M, v \models \varphi$, so by lemma 6.1: $M, v \models^{\mathcal{A}(w)} \varphi$, and therefore $M, w \models B\varphi$. ■

One of the corollaries of proposition 6.5 is that $\models_{mon\uparrow} L\varphi \Rightarrow \models_{mon\uparrow} A\varphi \rightarrow B\varphi$, which implies proposition 6.4, now for arbitrary unimodal φ, since $\models \varphi \Rightarrow \models L\varphi$ (**N**) is valid.

Corollary 6.1 *If $\varphi \in \mathcal{L}_{B,A,L}$ and $\models_{mon\uparrow} \varphi$, then $\models_{mon\uparrow} A\varphi \rightarrow B\varphi$.*

Notice that proposition 6.5 (and, *a fortiori* lemma 6.1) do not generalize to the many-agents case:

Example 6.2
Consider the two-agent singleton model $\langle \{w\}, \mathsf{R}_1, \mathsf{R}_2, \mathcal{A}_1, \mathcal{A}_2, V\rangle$ where $\mathsf{R}_1 = \mathsf{R}_2 = \{\langle w, w\rangle\}, \mathcal{A}_1(w) = \{p\}, \mathcal{A}_2(w) = \emptyset$ and V is arbitrary. Notice that the model trivially satisfies the structural and awareness requirements. Now let $\varphi = \neg B_2 p$. Then (i) $A_1\varphi = B_1(p \vee \neg p)$, so $w \models A_1\varphi$; (ii) $w \not\models^{\mathcal{A}_2(w)} p \Rightarrow w \not\models B_2 p \Rightarrow w \models \varphi \Rightarrow w \models L_1\varphi$; (iii) $w \not\mathrel{=\!|}^{\mathcal{A}_1(w)\cap\mathcal{A}_2(w)} p \Rightarrow w \not\mathrel{=\!|}^{\mathcal{A}_1(w)} B_2 p \Rightarrow w \not\models^{\mathcal{A}_1(w)} \varphi \Rightarrow w \not\models B_1\varphi$. In all, $w \models A_1\varphi \wedge L_1\varphi \wedge \neg B_1\varphi$.

Also notice that the converse of proposition 6.5 does not hold, irrespective of monotonicity conditions: it is easily verified that e.g. $\not\models B(p \vee q) \rightarrow B(p \vee \neg p)$.

However, there are other monotonicity results which are valid for the general multi-modal language, and are significant both from the perspective of correspondence theory, and from the perspective of doxastic application. These results are all related to the **4**- and converse **5**-axiom for belief in some way or other, and roughly amount to (*implicit*) introspection of *explicit* belief. First we state the correspondence results for the different monotonicity types:

Proposition 6.6

1. $B\varphi \to BB\varphi$ *is determined by mon↑ on frames.*
2. $\neg B\varphi \to L\neg B\varphi$ *is determined by mon↓.*
3. $\neg BB\varphi \to L\neg B\varphi$ *is determined by mon=.*

Proof:

1. First we show that the condition is sufficient. Let $\mathcal{A}$ be mon↑ with respect to R. Then $M, w \models B\varphi \Rightarrow$ for each u such that $w\mathsf{R}u$: $M, u \models^{\mathcal{A}(w)} \varphi \Rightarrow$ for each v and u such that $w\mathsf{R}v$ and $v\mathsf{R}u$: $M, u \models^{\mathcal{A}(w)\cap\mathcal{A}(v)} \varphi \Rightarrow M, w \models BB\varphi$. Necessity follows by contraposition: assume there is a non-mon↑ frame $\langle W, \mathsf{R}, \mathcal{A}\rangle$. Then there exist $w, v \in W$ and $p \in Prop$ such that $w\mathsf{R}v$ and $p \in \mathcal{A}(w) - \mathcal{A}(v)$. Then $w \models B(p \vee \neg p)$, yet $w \not\models BB(p \vee \neg p)$. Together this shows full correspondence.
2. To show that *mon↓* is sufficient, we argue indirectly. Let M be a *mon↓* model and suppose $w \models \neg B\varphi \wedge \neg L\neg B\varphi$. Thus for some v and u such that $w\mathsf{R}v$ and $w\mathsf{R}u$ and every v' for which $u\mathsf{R}v'$: (i) $v \not\models^{\mathcal{A}(w)} \varphi$ and (ii) $v' \models^{\mathcal{A}(u)} \varphi$. By euclidicity of R we have $u\mathsf{R}v$ and so (iii) $v \models^{\mathcal{A}(u)} \varphi$. Proposition 6.2, (iii) and *mon↓* imply (iv) $v \models^{\mathcal{A}(w)} \varphi$, contradicting (i). *mon↓* is also necessary, for if it does not hold for some frame we have worlds w and v such that $w\mathsf{R}v$ and a proposition p such that $p \in \mathcal{A}(v) - \mathcal{A}(w)$. Then however $w \models \neg B(p \vee \neg p) \wedge \neg L\neg B(p \vee \neg p)$.
3. sufficiency follows from (1) and (2), and necessity is proved analogously. ∎

The logics of special and general awareness partly agree with respect to their monotonicity behaviour. Compared to proposition 6.1, we notice that the validities in the first two lines still hold:

Proposition 6.7 *The implications displayed below are valid under mon↑, their converses under mon↓, and their bidirectional counterparts (i.e. equivalences) under mon=:*

$$\begin{array}{ll} \models_{mon\uparrow} A\varphi \to LA\varphi & \models_{mon\uparrow} L\neg A\varphi \to \neg A\varphi \\ \models_{mon\uparrow} B\varphi \to LB\varphi & \models_{mon\uparrow} L\neg B\varphi \to \neg B\varphi \end{array}$$

Proof: The assertions on the first line basically follow from the fact that $v \models A\varphi \Leftrightarrow \forall p$ in φ : $p \in \mathcal{A}(v)$; those on the second line follow from proposition 6.6. ∎

Despite this considerable overlap differences abound: $\models_{mon\uparrow} B\varphi \to BB\varphi$ holds in SAL, but not in GAL, and the reverse situation pops up for $\models_{mon\uparrow} B\neg B\varphi \to A\neg B\varphi$.

Inspection of the **5**-schema for explicit belief ($\neg B\varphi \to B\neg B\varphi$) shows that this requires a very strong condition on the awareness functions, in fact one which makes the system collapse: the distinction between B and L becomes vacuous. The situation here is very different from that in GAL, where at least the weakened form of the negative introspection axiom $(\neg B\varphi \wedge A\neg B\varphi) \rightarrow B\neg B\varphi$, could be obtained without running into a collapse.

Proposition 6.8

1. *$\neg B\varphi \rightarrow B\neg B\varphi$ is determined by overall totality of $\mathcal{A}$.*[30]
2. $\models_{total\ \mathcal{A}} B\varphi \leftrightarrow L\varphi$.

discussion

Now as for empirical adequacy, note that LO of types **N**, **I** and **E** is circumvented, but that **K** holds again. This means that agents are supposed to be entirely consistent in their beliefs, which is an idealization, of course. In some respects this logic is similar to the form of GAL where the awareness sets are closed under subformulas. General awareness as generated from a set of atoms seems even closer. F&H note however that the formula $B\varphi \rightarrow B(\varphi \vee \psi)$ is not valid according to this kind of general awareness, which is an advantage of this logic over that of special awareness. What is the precise relation between the logics of special and general awareness? To make the general comparison work (i.e. in the absence of monotonicity constraints), A_i has to be left out of the language. Then 'special' explicit beliefs can simply be pushed into the 'general' awareness sets. More formally, for any SAL model $M = \langle W, \{\mathsf{R}_i\}_i, \{\mathcal{A}_i\}_i, V\rangle$ define an equivalent GAL model $M' = \langle W, \{\mathsf{R}_i\}_i, \{\mathcal{A}'_i\}_i, V\rangle$, where $\mathcal{A}'_i(w) = \{\varphi \mid M, w \models B_i\varphi\}$. Then $M, w \models \varphi \Leftrightarrow M', w \models \varphi$ is shown by a straightforward induction proof. So,

Proposition 6.9 (reduction of SAL models)
Every SAL model gives rise to an equivalent GAL model.

6.4 Rantala semantics

Hintikka and especially Rantala[31] have proposed the addition of nonstandard worlds which are (according to Hintikka) doxastically or epistemically accessible but logically impossible. These rather mysterious entities are somewhat clarified by Rantala who suggests that nonstandard worlds are arbitrary indices that do not encounter a validity test, yet can be arbitrarily 'filled' in some cases. With a slightly disturbing shift

[30] Different from the previous results, for more agents the system has to be homogeneous with respect to **5** to obtain correspondence. So, the validity of $\neg B_i\varphi \rightarrow B_i\neg B_i\varphi$ requires the totality of *all* the $\mathcal{A}_i$'s: $\forall i, w : \mathcal{A}_i(w) = Prop$. The proofs are easy exercises.

[31] *Vide* [Hi75], [Ra82a] and (in a very general form) [Ra82b].

in terminology, Rantala adopts the phrase *non-normal* worlds for these non-standard objects.[32]

A Rantala model for the modal language $\mathcal{L}_{\vec{\Box}}$ is of the form $\langle W, W^*, \vec{\mathsf{R}}, V\rangle$. Here W is a set of 'normal' worlds and W^* a set of 'non-normal' worlds; it will be convenient to put $U = W \cup W^*$. Then $\mathsf{R}_i \subseteq U \times U$ and $V : \mathcal{L} \times U \longrightarrow \{0,1\}$. The truth conditions for normal worlds w are standard-type — for the connectives they are recursively specified from the assignments to the atoms. The truth conditions for non-normal worlds (that may enter when modal formulas are evaluated) are free:[33] for example, both φ and $\neg\varphi$ may be true in a non-normal world w^*, but then neither should be false; also neither might be true, but then they are both false. Truth conditions now are constraints on proper valuations: (w is normal, u may be non-normal, i.e. $w \in W$ and $u \in U$)

- $V(\neg\varphi, w) = 1$ iff $V(\varphi, w) = 0$;
- $V(\varphi \wedge \psi, w) = 1$ iff $V(\varphi, w) = V(\psi, w) = 1$;
- $V(\Box_i\varphi, w) = 1$ iff $V(\varphi, u) = 1$ for each u such that $w\mathsf{R}_i u$;
- $\models_{\mathcal{C}} \varphi$ iff $V(\varphi, w) = 1$ for each model $\langle W, W^*, \vec{\mathsf{R}}, V\rangle \in \mathcal{C}$ and $w \in W$.

It can easily be shown that Rantala semantics is entirely flexible: every modal system that contains **pL** is characterized by a class of Rantala models.[34] As an example of the force of this framework [Wa90] shows that

Proposition 6.10 (Wansing)
Each GAL *model induces a globally equivalent Rantala model.*

Proof:[35] Given a GAL model $M = \langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$, let $M' = \langle W, W^*, \vec{\mathsf{B}}, \vec{\mathsf{R}}, \vec{\mathsf{A}}, V'\rangle$ be a structure such that $W^* = \{\mathcal{A}_i(w) \mid i \leq m \;\&\; w \in W\}$, $w\mathsf{A}_i v \Leftrightarrow v = \mathcal{A}_i(w) \;\&\; w \in W$, $\mathsf{B}_i = \mathsf{R}_i \cup \mathsf{A}_i$. B_i, L_i, A_i are interpreted by means of $\mathsf{B}_i, \mathsf{R}_i, \mathsf{A}_i$ respectively. $V'(p, w) = V(p, w)$ if $w \in W$, and $V'(\varphi, w^*) = 1$ iff $\varphi \in w^* \;\&\; w^* \in W^*$. Then imposing the usual truth conditions on normal worlds for connectives and modal operators turns the structure into a Rantala model that is equivalent to M on normal worlds, and therefore the two models are globally equivalent, i.e. verify the same formulas. ∎

[32] Rantala's normal worlds correspond to [Kr65b]'s *designated* worlds rather than to Kripke's normal worlds. With Kripke, checking validity takes place in designated worlds, with Rantala in *his* normal worlds. There is also a difference in the truth conditions for $\Box$: Kripke's non-normal worlds w^* reject any belief whatsoever, i.e. $w^* \not\models \Box\varphi$, whereas Rantala's non-normal worlds w^* allow an arbitrary truth assignment to beliefs, i.e. whether or not $w^* \models \Box\varphi$ is stipulated by the model. So with Rantala, **N** is eliminated not by some constraint on the starting point of evaluation, but by missing information in the accessible worlds.

[33] So, they are *not* open, i.e. the semantics is still total. However, in general truth values are not recursively specified.

[34] Cf. [PW89], [Wa89]; the canonical model and truth lemma are rather straightforward.

[35] Our proof slightly departs from Wansing's since he does not treat A_i as a genuine modal operator, with an accessibility relation A_i of its own.

6.5 Sieve models

Although proposition 6.10 demonstrates that Rantala models are well-equipped for awareness logics, it does not show their superiority, neither as a specific description of a certain type of awareness, nor as a general framework. The point is that a slight generalization of the GAL models already provides an equally basic and flexible framework: let us simply drop the structural conditions on R_i (i.e. do not require seriality, transitivity and euclidicity).[36] Technically, these *sieve models* are of the form $\langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$ where $\langle W, \mathsf{R}_i, V\rangle$ is a common Kripke model and $\mathcal{A}_i(w) \subseteq \mathcal{L}_{\vec{\Box}}$ for all $w \in W$. The essential truth condition is like the GAL one for explicit belief:

$$w \models \Box_i\varphi \quad \text{iff} \quad \varphi \in \mathcal{A}_i(w) \;\&\; v \models \varphi \text{ for all } v \text{ such that } w\mathsf{R}_i v$$

Then we can prove a converse to Wansing's result:

Proposition 6.11 (reduction of Rantala models to sieve models)
Every Rantala model induces a globally equivalent sieve model.

Proof: A Rantala model $M = \langle W, W^*, \vec{\mathsf{R}}, V\rangle$ can be transformed into a sieve model $M' = \langle W, \vec{\mathsf{R}}', \vec{\mathcal{A}}, V'\rangle$, by taking[37]

- $\mathsf{R}_i' = \mathsf{R}_i \cap W \times W$
- $\mathcal{A}_i(w) = \{\psi \mid V(\psi, v) = 1 \text{ for all } v \in W^* \text{ such that } w\mathsf{R}_i v\}$
- $V'(p, w) = V(p, w)$ for all $p \in Prop, w \in W$.

A straightforward induction on the structure of the formulas shows simulation of truth on these models:

$$V(\varphi, w) = 1 \Leftrightarrow M', w \models \varphi, \text{ for all } w \in W.$$

We will prove the induction step for $\Box_i$, the other steps are omitted: assume the assertion to hold for some φ (IH), then $V(\Box_i\varphi, w) = 1 \Leftrightarrow \forall v \in \mathsf{R}_i[w] : V(\varphi, v) = 1 \Leftrightarrow \forall v \in \mathsf{R}_i[w] \cap W : V(\varphi, v) = 1 \;\&\; \forall v \in \mathsf{R}_i[w] \cap W^* : V(\varphi, v) = 1 \Leftrightarrow$ (by IH + defs. R_i' and $\mathcal{A}_i$) $\forall v \in \mathsf{R}_i'[w] : M', v \models \varphi \;\&\; \varphi \in \mathcal{A}_i(w) \Leftrightarrow M', w \models \Box_i\varphi$.
Therefore, since global truth is restricted to W, $M \models \varphi \Leftrightarrow M' \models \varphi$. ■

The last two propositions show that Rantala's non-normal world semantics and sieve semantics are equivalent, and therefore equally flexible (the structural conditions for GAL models do not interfere in Wansing's result). Combining the previous

[36] After reading an early version of [Th91a], Joe Halpern sent me the chapter 'Dealing with logical omniscience' of a forthcoming book on *Reasoning about Knowledge*, which he is writing in collaboration with Ron Fagin, Yoram Moses and Moshe Vardi. The section on awareness shows the same relaxation of structural constraints, which, like with us, may have been prompted by Wansing's embedding of GAL models in non-normal world semantics (proposition 6.10). Consequently, Halpern's chapter and the present chapter share a number of results, such as proposition 6.11.

[37] Almost the same construction has independently been suggested by Halpern, cf. [Wa90, note 7]. The generality of the result was apparently overlooked, but Halpern [pers. comm.] notes that this is how he intended his comment on Wansing's manuscript to be interpreted.

remarks we obtain a corollary, which is useful for deriving more specific completeness theorems. It can also be shown directly.

Corollary 6.2 *Every modal system containing* **pL** *is characterized by a canonical sieve model.*

Direct proof: For a modal system $\mathbf{S} \supseteq \mathbf{pL}$, define the canonical model $\mathcal{M} = \langle \mathcal{W}, \vec{\mathcal{R}}, \vec{\mathcal{A}}, \mathcal{V} \rangle$ by:

- $\mathcal{W}$ is the set of maximally **S**-consistent sets of formulas,
- $\Gamma \mathcal{R}_i \Delta \quad$ iff $\quad \Box_i^{-1}[\Gamma] \subseteq \Delta$,
- $\mathcal{A}_i(\Gamma) = \Box_i^{-1}[\Gamma]$,
- $\mathcal{V}(p, \Sigma) = 1 \quad$ iff $\quad p \in \Sigma$.

This enables an easy proof of the truth lemma; since **S** contains **pL** we also have Lindenbaum's lemma. Together this shows completeness in the usual way. ■

A first grumbling remark here is that the proof is almost too easy; the reason for this ease is that the $\mathcal{A}$-sets allow an enormous amount of freedom. Specific logics will be more difficult to handle, since we will be inclined to impose the constraint on the R-relations instead of on the $\mathcal{A}$-functions, which is impossible in the general case. Also in this respect sieve semantics and Rantala's non-normal world semantics are comparable: in the latter case the *valuation* type is the second dimension of freedom. This comparison also indicates that it is somewhat dubious to count $\mathcal{A}$ as part of the *frame*, as we did earlier. A second remark is that for *arbitrary* multi-modal logics, the R in the sieve models can be dismissed, turning the semantics into 'syntax in disguise'. So, for the homogeneous modal language $\mathcal{L}_{\vec{\Box}}$ the completeness proof above can be simplified by taking $\mathcal{R}_i = \emptyset$, but the earlier clause fits the heterogeneous language $\mathcal{L}_{\vec{L},\vec{A},\vec{B}}$, where the logic for each L is normal and $B\varphi \Leftrightarrow L\varphi \wedge A\varphi$. So, for awareness logics the coexistence of explicit and implicit belief makes us keep the R after all.

6.6 Neighbourhood semantics

In this section we review neighbourhood semantics as a framework for providing awareness logic. Then we will introduce cluster models and show how they are related to neighbourhood structures, and we will argue that the latter are still preferable. A similar story can be told for Jaspars' construal of the [RB80] fusion models. Finally we will argue that there is no technical reason to maintain such models in the presence of the framework of sieve models.

Neighbourhood or 'Scott-Montague' (SM) semantics can be regarded as a topological or functional generalization of Kripke semantics.[38] First consider a classical Kripke model $M = \langle W, \mathsf{R}, V \rangle$. We employ some abbreviations:

- $[\![\varphi]\!]_M = \{v \mid M, v \models \varphi\}$;

[38] [Mo68], [Sc70]; the functional view has been elaborated by David Lewis.

- $\mathsf{R}[w] = \{v \mid w\mathsf{R}v\}$.

The truth condition for necessity can thus be reformulated as:

$$w \in [\![\Box\varphi]\!] \quad \text{iff} \quad \mathsf{R}[w] \subseteq [\![\varphi]\!],$$

in other words, iff $[\![\varphi]\!] \in \{X \subseteq W \mid \mathsf{R}[w] \subseteq X\}$. From this it is but a small step to replace the principal filter $\{X \subseteq W \mid \mathsf{R}[w] \subseteq X\}$ by an arbitrary set of subsets of W, a 'neighbourhood' (of w). So an SM model $\langle W, \mathsf{N}, V\rangle$ has the usual W and V and $\mathsf{N}(w) \subseteq \wp(W)$. The key truth condition is:

$$w \models \Box\varphi \quad \text{iff} \quad [\![\varphi]\!] \in \mathsf{N}(w)$$

The neighbourhood metaphor is clearly inspired by topology (abstract geometry) and this way of putting things certainly has pictorial advantages. However, for symbolic manipulation another formulation is more apt. Compared to traditional mathematics, the alternative format is closer to algebra than to geometry. The idea is to consider (syntactic) modal operators also as semantic operators. In a functional model $\langle W, f, V\rangle$ a modal $\Box$ is interpreted by an operator $f : \wp(W) \longrightarrow \wp(W)$, i.e.

$$[\![\Box\varphi]\!] = f([\![\varphi]\!]).$$

These are just two ways of saying the same thing, though.[39]

Proposition 6.12
Neighbourhood semantics and functional semantics are equivalent.

Proof: notice that f and N are interdefinable: $w \in f(X) \Leftrightarrow X \in \mathsf{N}(w)$ ■

completeness and correspondence

A completeness theorem is easily found and proven: [40]

Theorem 6.3 (Segerberg) *The modal logic for* SM *semantics is* **pL + E**.

So, omniscience of types **N**, **K**, **C** or **I** can be avoided. Moreover, SM semantics is of considerable flexibility, as can be seen in a number of correspondences:

C: (*intersection*) $f(X) \cap f(Y) \subseteq f(X \cap Y)$

N: (*fixed unit*) $f(W) = W$

D: (*consistency*) $f(X) \cap f(W - X) = \emptyset$

[39] This duality is reminiscent of the situation in generalized quantifier theory, where one encounters a functional vs. a relational view on interpreted determiners, comparable to our neighbourhood vs operator view.

[40] See [Ch80].

4: (*interior-property*) $f(X) \subseteq f(f(X))$

5: (*exterior-property*) $W - f(X) \subseteq f(W - f(X))$

Now interpreting $\Box$ as *explicit belief* (B), which requirements should be imposed on **N** or f? As a prerequisite to this, observe the following evident postulates for B:

D* $\vdash \neg B(\varphi \wedge \neg\varphi)$

$\mathbf{C_c}$ $\vdash B(\varphi \wedge \psi) \rightarrow (B\varphi \wedge B\psi)$

These principles have simple semantic counterparts:

D*: (*fixed zero*) $f(\emptyset) = \emptyset$

$\mathbf{C_c}$: (*upward monotonicity*) $X \subseteq Y \Rightarrow f(X) \subseteq f(Y)$

Accepting $\mathbf{C_c}$ implies accepting **I** too, because of the validity of **E**. So we are stuck with some types of LO after all. As a matter of fact, it seems that **E** and **I** are only slightly weaker than **N**: assume[41] some simple observation p and some complex mathematical truth φ. So $\models \varphi$ and consequently $\models p \leftrightarrow (p \wedge \varphi)$, and therefore $\models Bp \leftrightarrow B(p \wedge \varphi)$. In words, if one believes some fact, then one will also believe that fact and a complicated piece of mathematics. We may conclude that for these severe types of LO neighbourhood semantics is no great help, but it may be used for some types of omniscience, just like the following two systems do.

6.6.1 Cluster models

To motivate yet another logic, F&H claim that

> Although the logic of general awareness is quite flexible, it still has the property that an agent cannot hold inconsistent beliefs. [...] Our key observation is that one reason that people can hold inconsistent beliefs is that beliefs tend to come in non-interactive clusters. [FH88, p.58]

In this respect an agent is similar to a community in which different persons may have different opinions, yet no one will defend contradictions. In a nutshell, beliefs stemming from various frames of mind need not be combined by the agent. In particular, we may want $B\varphi \wedge B\neg\varphi$ to be satisfiable, but $B(\varphi \wedge \neg\varphi)$ not. So clearly axiom schema **C** has to be rejected for this logic. To this purpose F&H propose what we will call 'cluster models'.[42]

[41] This argument was suggested by René Ahn.

[42] In [FH88, sect.6] cluster models are called 'Kripke structures for local reasoning'.

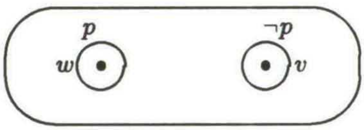
semantics

Cluster models are of the form $\langle W, \{\mathsf{C}_i\}_i, V\rangle$, where W and V are as usual, $\mathsf{C}_i(w) \subseteq \wp(W) - \{\emptyset\}$ and $\mathsf{C}_i(w) \neq \emptyset$ for each $w \in W$. So $\mathsf{C}_i(w)$ is a nonempty set of nonempty sets of worlds. The truth conditions for the connectives are standard-type and those for the doxastic operators run as follows:[43]

- $M, w \models L_i\varphi$ iff $M, v \models \varphi$ for every v such that $v \in \bigcap_{T\in \mathsf{C}_i(w)} T$ (i.e. $\bigcap \mathsf{C}_i(w) \subseteq [\![\varphi]\!]$);
- $M, w \models B_i\varphi$ iff $M, v \models \varphi$ for all v in some particular $T \in \mathsf{C}_i(w)$ (i.e. there is a $T \in \mathsf{C}_i(w)$ such that $T \subseteq [\![\varphi]\!]$, or: $\mathsf{C}_i(w) \cap \wp[\![\varphi]\!] \neq \emptyset$).

These clauses enable an evaluation of the behaviour of explicit belief with respect to the different sorts of omniscience: **N**-omniscience is obviously restored, i.e. **N** holds for explicit belief once again. For if φ is valid, then $[\![\varphi]\!] = W$ for any model, and so $B_i\varphi$ is always true. **I** is also easily proved valid by the transitivity of $\subseteq$. **C** is eliminated, however:

Example 6.3 *A simple counter-example for* $(Bp \wedge B\neg p) \rightarrow B(p \wedge \neg p)$ *is:*



Here $\mathsf{C}(w) = \{\{w\}, \{v\}\}$.

This implies that **K**-omniscience is avoided too, since **C** and **K** are deductively equivalent, modulo **I** and the propositional calculus. The above countermodel shows that the 'consistency' axiom **D**, which is equivalent to $\neg(B\varphi \wedge B\neg\varphi)$, is also invalid. Yet the (by **I**) weaker axiom **D*** is validated: (for B, not for L !)

D* $\qquad \vdash \neg B(\varphi \wedge \neg\varphi)$,

i.e. $\vdash \neg B\bot$. We are ready for a completeness result.

Theorem 6.4 (Fagin&Halpern) [44]
The modal system for the logic of local reasoning is **K** *(i.e.* **NK***) for* L_i, **NID*** *for* B_i, *together with the connecting axiom* $\vdash B_i\varphi \rightarrow L_i\varphi$.

A proof of this fact and a discussion of extensions of the system along the dimension of introspection is omitted since we can give another, very rewarding result which obviates the completeness theorem. This is achieved by a direct correspondence between cluster models and SM-models. Since truth of explicit belief amounts to *containment* of a set of the relevant cluster, and to *membership* of the relevant neighbourhood, we at least have to require the neighbourhoods to be increasing (=upwards monotone).

[43]In an earlier version of [FH88], presented on IJCAI85, another operator S_i ('strong belief') pops up, with interpretation $M, w \models S_i\varphi$ iff $M, v \models \varphi$ for all $T \in \mathsf{C}_i(w)$ and $v \in T$ ($\Leftrightarrow \bigcup \mathsf{C}_i(w) \subseteq [\![\varphi]\!]_M$).

[44][FH88, theorem 8.5]

Proposition 6.13 (reduction of cluster to neighbourhood models)
For explicit belief, cluster models are equivalent to **monotonically increasing** *neighbourhood structures with* **fixed unit** *and* **fixed zero**. *For implicit belief the related neighbourhoods are in addition* **intersective** *(and therefore* **filters***).*

Proof: Given a cluster model $\langle W, \mathsf{C}_i, V\rangle$ one can easily construct a neighbourhood structure for B_i by adding supersets to the clusters:

$$\mathsf{N}_i(w) = \mathsf{C}_i^{\uparrow}(w) = \{X \mid T \subseteq X \text{ for some } T \in \mathsf{C}_i(w)\}$$

N_i is increasing and non-trivial, i.e. $\mathsf{N}_i(w) \neq \emptyset$ and $\mathsf{N}_i(w) \neq \wp(W)$, since $\mathsf{C}_i(w) \neq \emptyset$ and $\emptyset \notin \mathsf{C}_i(w)$. The neighbourhoods for L_i are formed by taking intersections:

$$\mathsf{N}_i'(w) = \{X \mid \bigcap \mathsf{C}_i(w) \subseteq X\}$$

Then $\mathsf{N}_i'(w)$ is either a principal filter or degenerated into $\mathsf{N}_i'(w) = \wp(W)$.

Now if B_i is interpreted by N_i and L_i by N_i' in the model $\langle W, \mathsf{N}_i, \mathsf{N}_i', V\rangle$ (i.e. $w \models B_i\varphi \Leftrightarrow [\![\varphi]\!] \in \mathsf{N}_i(w)$, etcetera), then a straightforward induction shows that both models are equivalent (i.e. verify the same formulas in the same worlds).

To show the other direction, assume a neighbourhood model $\langle W, \mathsf{N}_i, \mathsf{N}_i', V\rangle$ where N_i is non-trivial and increasing, and N_i' is its closure under arbitrary intersections and supersets. In fact there exist several correct choices for related clusters:

- $\mathsf{C}_i = \mathsf{N}_i$;
- $\mathsf{C}_i(w) = \mathsf{N}_i^{\downarrow}(w) = \{X \in \mathsf{N}_i(w) \mid Y \subset X \text{ for no } Y \in \mathsf{N}_i(w)\}$ (the $\subseteq$-minimal elements of N_i)

Again an inductive argument proves equivalence of the cluster model and the SM model. ∎

Theorem 6.4 now follows as an almost immediate corollary.

The obvious advantage of such a reduction is that a lot of results become available. To wit, for neighbourhood structures **4** holds precisely on those neighbourhood frames that have the interior-property, which after translation into cluster semantics yields:

$$X \in \mathsf{C}_i^{\uparrow}(w) \Rightarrow \{v \mid X \in \mathsf{C}_i^{\uparrow}(v)\} \in \mathsf{C}_i^{\uparrow}(w)$$

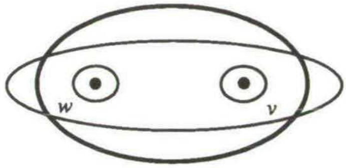
F&H propose a different condition to ensure positive introspection for both types of belief:

$$v \in T \in \mathsf{C}_i(w) \Rightarrow T \in \mathsf{C}_i(v)$$

This elegant condition implies interiority, and would be preferable because of its simplicity. Unfortunately it is too strong, i.e. it verifies **4** but does not correspond to it. Here is a counter-example to full correspondence:

Example 6.4
Let $W = \{w, v\}$, $\mathsf{C}(w) = \{\{w\}, \{v\}, \{w, v\}\}$ and $\mathsf{C}(v) = \{\{w, v\}\}$. See figure 6.1 (in the diagrams, $\mathsf{C}(w)$ is indicated by thin lines, $\mathsf{C}(v)$ by thick lines). $\mathsf{C} = \mathsf{C}^{\uparrow}$ has the interior-property and therefore verifies **4**. *However it does not conform to the above condition: $v \in \{v\} \in \mathsf{C}(w)$, but $\{v\} \notin \mathsf{C}(v)$.*

Figure 6.1:



A very similar story can be told for negative introspection. **5** schemata are verified on frames having the exterior-property:

$$X \notin \mathsf{C}_i^{\uparrow}(w) \Rightarrow \{v \mid X \notin \mathsf{C}_i^{\uparrow}(v)\} \in \mathsf{C}_i^{\uparrow}(w)$$

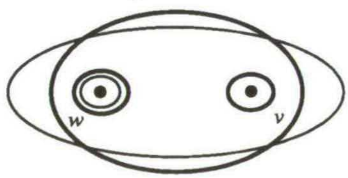
F&H again propose a much more simple condition:

$$v \in T \in \mathsf{C}_i(w) \Rightarrow \mathsf{C}_i(v) \subseteq \mathsf{C}_i(w)$$

Again this condition is sufficient but not necessary witness the following counter-example:

Example 6.5
Let $W = \{w, v\}$, $\mathsf{C}(w) = \{\{w\}, \{w, v\}\}$ *and* $\mathsf{C}(v) = \{\{w\}, \{v\}, \{w, v\}\}$, *see figure 6.2.*

Figure 6.2:



$\mathsf{C} = \mathsf{C}^{\uparrow}$ *has the exterior-property and thus verifies* $\neg B\varphi \rightarrow B\neg B\varphi$. *However it does not conform to the above condition:* $v \in \{w, v\} \in \mathsf{C}(w)$ *and* $\mathsf{C}(v) \not\subseteq \mathsf{C}(w)$.

6.6.2 Fusion models

[Ja91b] deals with 'confused' belief, as he calls it, suggesting to solve the problem of why incompatible beliefs apparently do not lead to total mental collapse. To this purpose Jaspars uses the idea of 'fusion' of worlds from [RB80], but without the need to stipulate non-standard worlds created by algebraic operations on ordinary worlds.[45]

[45]Cf. [Va86] for a different though similar implementation of the 'fusion' idea of [RB80]; Vardi's account is closer to the original idea, where fusion is achieved by lattice-like operations on worlds.

In a fusion model $M = \langle W, \mathsf{R}_i, V\rangle$ an accessibility relation $\mathsf{R}_i \subseteq W \times (\wp(W) - \{\emptyset\})$ typically connects worlds to 'fused' sets of worlds instead of to single worlds. The crucial truth condition is essentially:

- $M, w \models B_i\varphi$ iff $X \cap [\![\varphi]\!]_M \neq \emptyset$ for all X such that $w\mathsf{R}_iX$.

Where [RB80] adds conditions to normalize the logical system, Jaspars considers the pure semantics and demonstrates its soundness and completeness for the modal system multi-**NI**. Our point is that the same strategy applied to cluster models can be used here too: a reduction to neighbourhood models is feasible. This may be surprising given the 'second order' nature of the above truth condition.

Proposition 6.14 (reduction of fusion models)
Fusion models correspond to **strong, monotonically increasing** *neighbourhood structures.*

Proof: Starting with a fusion model $M = \langle W, \mathsf{R}_i, V\rangle$, one can construct a neighbourhood structure $N = \langle W, \mathsf{N}_i, V\rangle$ for B_i by:

$$\mathsf{N}_i(w) = \{X \mid \forall Y : w\mathsf{R}_iY \Rightarrow X \cap Y \neq \emptyset\}.$$

N_i is obviously increasing and strong; by induction M and N are equivalent.

For the other direction, let $N = \langle W, \mathsf{N}_i, V\rangle$ be a neighbourhood model in which every N_i is strong and increasing. Then a related fusion model $M = \langle W, \mathsf{R}_i, V\rangle$ can be defined:

$$w\mathsf{R}_iX \Leftrightarrow X \subseteq W \text{ and } X \cap Y \neq \emptyset \text{ for all } Y \text{ such that } Y \in \mathsf{N}_i(w).$$

This definition is correct since $w\mathsf{R}_iX \ \&\ W \in \mathsf{N}_i(w) \Rightarrow \emptyset \neq X \subseteq W$. We prove the key induction step for the equivalence of the models, assuming $[\![\varphi]\!]_M = [\![\varphi]\!]_N$ (IH):
$M, w \models B_i\varphi \Leftrightarrow$ for all X such that $w\mathsf{R}_iX$: $X \cap [\![\varphi]\!]_M \neq \emptyset \Leftrightarrow$ (IH) for each X such that $X \cap Y \neq \emptyset$ for all $Y \in \mathsf{N}_i(w)$: $X \cap [\![\varphi]\!]_N \neq \emptyset \Leftrightarrow^* [\![\varphi]\!]_N \in \mathsf{N}_i(w) \Leftrightarrow N, w \models B_i\varphi$.
Here $\Leftarrow^*$ is obvious (take $Y = [\![\varphi]\!]$) and $\Rightarrow^*$ follows by an indirect argument: suppose that $[\![\varphi]\!] \notin \mathsf{N}_i(w)$ and for each X such that $X \cap Y \neq \emptyset$ for all $Y \in \mathsf{N}_i(w)$: $X \cap [\![\varphi]\!] \neq \emptyset$. Since N_i is monotonically increasing $Y \not\subseteq [\![\varphi]\!]$ for all $Y \in \mathsf{N}_i(w)$, so for all $Y \in \mathsf{N}_i(w)$: $Y \cap [\![\varphi]\!]^c \neq \emptyset$. The choice $X = [\![\varphi]\!]^c$ thus leads to the contradiction $[\![\varphi]\!] \cap [\![\varphi]\!]^c \neq \emptyset$. So $[\![B_i\varphi]\!]_M = [\![B_i\varphi]\!]_N$. ∎

The procedure used in this proof again provides an effective way to incorporate additional axioms, such as **D*** and **4**. The corresponding conditions can thus be derived and coincide to those presented in [Ja91b].[46]

[46]To wit, the first transformation used in the proof gives that $X \notin \mathsf{N}(w) \Leftrightarrow w\mathsf{R}^\uparrow X^c$. Thus the **D*** condition $\emptyset \notin \mathsf{N}(w)$ amounts to $w\mathsf{R}^\uparrow W$. The usual **4** condition translates after contraposition and replacement of X^c by Y to $w\mathsf{R}^\uparrow\{v \mid v\mathsf{R}^\uparrow Y\} \Rightarrow w\mathsf{R}^\uparrow Y$.

6.6.3 comparison

Given the reductions to SM models we see no compelling reason to create a new kind of semantics. The somewhat greater intuitive appeal of the clusters which are generally smaller than neighbourhoods is nullified by the difficulty in formulating structural constraints, as illustrated above. And furthermore, neighbourhood models are based on the simple idea that a proposition corresponds to the set of worlds in which it is true. Therefore, SM semantics cannot distinguish logically equivalent propositions, which is generally considered unacceptable for awareness logics.

Moreover, it is doubtful whether these models should coexist with those for general awareness. It seems that F&H have overlooked the fact that cluster models essentially form a special case of the general models for awareness: analogously to the earlier argument for containment of special awareness into general awareness structures, we can reduce cluster models to general awareness models by simulating validated explicit beliefs in the awareness sets.

Proposition 6.15 (reduction of cluster to sieve models)
Every cluster model induces an equivalent sieve model.

Proof:[47] A cluster model $\langle W, \vec{\mathsf{C}}, V\rangle$ can be transformed into a sieve model $M' = \langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$, by taking

- $w\mathsf{R}_i v \quad \text{iff} \quad v \in \bigcap \mathsf{C}_i(w)$,
- $\mathcal{A}_i(w) = \{\psi \mid M, w \models B_i\psi\}$.

A straightforward induction shows simulation of truth on models:

$$M, w \models \varphi \quad \text{iff} \quad M', w \models \varphi$$

■

The proviso is here that we must be willing to give up the structural conditions on accessibility. So, seriality, transitivity and euclidicity have to be eliminated; but the same holds for the treatment of implicit belief in cluster models. More in particular, we believe there is little motivation to superimpose the awareness function to cluster models or the like, as [FH88, p.61] propose.

A similar and in some respect more general reduction of SM models to sieve models can also be obtained, by simply taking $\mathsf{R}_i = \emptyset$ in the above proof.

Proposition 6.16 (reduction of SM to sieve models)
Every neighbourhood structure induces an equivalent sieve model.

Despite the fact that neighbourhood semantics is usually believed to be the weakest modal semantics, providing a universal modal framework, we notice that both Rantala and sieve models are more general. Therefore, the converse of the last proposition does not hold, as follows from the respective completeness theorems.

---

[47] The referee of [Th91a] rightly notices that this proposition also follows from [Wa90, claim 4] and proposition 6.11 here.

6.7 Conclusion and afterthoughts

We have studied a number of 'total' awareness logics based on modifications of classical possible world semantics, either Kripke-style or Scott/Montague-style. Most of the research reported here was more or less directly connected to the logics proposed in [FH88]. In retrospect, our contribution has two different aspects: one is theory-internal and one theory-external.

On the one hand we have developed some total awareness logics further, obtaining detailed results concerning completeness and correspondence. On the semantic side conditions of *monotonicity*, meaning roughly that what one is aware of in some world will still be present in the alternatives considered possible, turned out to be important. Monotonicity conditions often corresponded to (weak) introspection properties of explicit or actual belief. We notice that though we have solved these correspondence problems, there are a number of open questions, in particular related to complete and natural axiomatizations of validity in different (sub)languages.

On the other hand we have compared these logics along the dimensions of generality and flexibility. Slightly generalizing F&H's logic of general awareness GAL, we obtained a fully general and flexible logic containing what we called a 'sieve semantics', which was shown to be equivalent to Rantala's (im)possible world semantics. Just as [Wa90] embedded the logics proposed in [FH88] in Rantala semantics, we were able to embed them in sieve semantics. In particular, we may summarize the relations between these logics as follows:

- The most general frameworks capable of modelling every logic containing the classical tautologies are sieve semantics and Rantala's non-normal world semantics.

- Subordinated to these most general frameworks are neighbourhood semantics and the semantics of GAL (which amounts to sieve semantics with accessibility conditions). These two logics are incomparable: for example, **E** holds in neighbourhood structures, but not hold in GAL, **D** is valid in GAL but generally not in neighbourhood semantics.

- The special awareness logic is a special case of GAL, i.e. it is stronger than GAL in the sense that it validates more.

- Cluster and fusion models may be considered to be special cases of neighbourhood structures; their logics are therefore stronger.[48]

Despite the comfort of these results, there are some worries.

Is the most general (awareness) logic necessarily the best? Although 'everything' can be expressed in generalized awareness logic, this logic may not always be preferable. A lot depends on the particular application one has in mind. As Halpern[49]

[48] In fact the completeness properties indicate that fusion models are more general than cluster models, since the former validate less.

[49] Private communication, July 1991

notices, resource-bounded reasoning can be modelled quite naturally within sieve semantics, but it is awkward at best to think of it as a special case of Rantala semantics. Also, despite the text reduction, the type of awareness connected to varying belief in different frames of mind may be modelled more naturally in cluster semantics. This, of course, is largely a matter of intuition. But then again, how plausible are the sieve models? After all, the sieve models are based on the notion of possible worlds, which is an abstraction, especially in the context of belief and knowledge. So, one may argue that replacing total worlds by *partial worlds* (or, situations) will be a clear improvement. Moreover, combined with a suitable notion of validity, partial semantics will automatically exclude a lot of logical omniscience. We will make this move in the next chapter, but, especially for a logic with both explicit and implicit belief, sieve semantics is concise and convenient.

Is the generalized awareness logic obtained by sieve semantics still a logic in the usual sense of the word? In fact, this logic is classical in the sense that it contains all propositional tautologies, but it is certainly non-classical in the sense that it can model every consistent set of formulas including these tautologies as if it were a logic. To some this may go beyond what might be called a 'logic' proper, yet to us this seems to be an inevitable consequence of modelling such psychological notions as (actual) belief or knowledge.

Chapter 7

Partial logics of awareness

In the previous chapter we considered total approaches to awareness and explicit belief, i.e. bivalent varieties of possible world semantics. Here we advocate a *partial* approach to conscious belief. In fact, similar to the richness noticed in part one, there are several partial approaches to awareness.

7.1 Introduction and overview

Is there a need for a renewed look at awareness? In the previous chapter we showed that there are very powerful frameworks (viz. that of *sieve* semantics and *non-normal* world semantics) that can solve the problem of modelling an arbitrary modal logic that extends the classical propositional calculus. Although the problem of modelling weak logics for such psychological notions as awareness is thus solved on a technical level, one often would prefer a more natural representation device.

A more natural approach to the virtues of awareness and the vices of logical omniscience (LO) is to move to partial semantics, where the classical truth values (viz. *true* and *false*) may be *undefined* and sometimes even *overdefined*, leading to an, essentially, 3- or 4-valued logic. After all, the very notion of partiality was motivated by the idea that one conceives or considers only part of the world, i.e. the part one is aware of in one's perception or belief. Therefore we proceed by reinspecting a standard approach to partial modal logic in the next section and investigate into its suitability for modelling belief and awareness.

Some of the expressive deficiency of the standard partial approach can be removed by adding new connectives, to wit: non-standard *negation* and *implication*, while keeping the semantics for the old connectives straight. Since this gives rise to renewed omniscience we also study another approach: keep the syntax straight, but allow a dual perspective on truth, both partial and complete, within a single system. We also discuss ways of dealing with residual problems of awareness, and compare these with each other, and with total approaches. For reasons of space and preference, we restrict ourselves to coherent (three-valued) models here.[1]

[1] [Th92b] reviews some four-valued approaches advocated by Levesque and others.

7.2 A purely partial approach

The initial logical representation language is that of multi-modal propositional logic $\mathcal{L}_{\neg,\wedge,\{B_i\}_i}(Prop)$ ($\mathcal{L}_{\vec{B}}$ for short), with operators B_i standing for 'agent i explicitly believes that'. Its semantics consists of the *coherent* models with the standard truth/falsifity clauses given in section 4.3. The consequence notion will be that of relative verification.

Because there are *no* valid formulas in this form of partial semantics, the usual types of omniscience connected to the modal schemes[2] **K** and **C** are circumvented. Moreover, though the inference rules **N**, **I** and **E** are vacuously valid (since the validity of the premise cannot be realized), they are innocuous now: these rules have no input, and therefore no output either. For example, $B(p \vee \neg p)$ and $B(Bp \vee \neg Bp)$ are neither valid nor produced by **N**.

Though the logic deals with belief rather than with awareness, it also provides an indirect route to awareness: recall from section 6.3.2 that somebody may be said to be aware of (or, 'acquainted with') φ, if every primitive p in φ has a definite truth value (1 or 0) in every situation the agent considers possible from the situation she is in, in other words, if she explicitly believes $p \vee \neg p$.[3] Deriving awareness from explicit beliefs is a promising way to reintroduce one of the central notions in the field.

Since the usual types of LO are circumvented, and awareness can be derived, the purely partial semantics for this multi-modal logic may seem quite successful, and this could be the end of the story. However, there are a number of complications:

- the impossibility of absolute validity apparently excludes the incorporation of additional properties which are needed to model various types of knowledge and belief. Positive and negative introspection ('knowing of what we (do not) know that we (do not) know it'), truth of knowledge, and consistency of belief ('not believing contradictions') cannot be encoded in the usual schemes **4**, **5**, **T** and **D**, respectively. This will prove to be a minor point.

- the impossibility of absolute validity also excludes intuitively correct *objective* truths such as $Bp \vee \neg Bp$.[4] More generally, one would prefer a logic that at least contains (the modal substitutions of) classical propositional logic. This is a major point.

- unlike absolute validity, we do obtain relative validity. Then it turns out that many of the eliminated forms of LO pop up again in relativized form. This is also a major point.

In a way, the first point is cancelled by the third: if the usual types of LO, which are captured by basic modal schemes, are obtainable in a relative shape, this may

[2]See the introduction to chapter 6.

[3]See the definition of $A\varphi$ on page 161.

[4]This may be contrasted to a *subjective* truth such as $B(p \vee \neg p)$. I use the terms objective/subjective in an intuitive sense. The distinction involved does not correspond to non-modal vs fully modalized (cf. [Le84a] etcetera), but to 'intuitively valid' vs 'intuitively contingent', we suppose.

also hold for schemes such as **4**. For, as in the well-known deduction theorem of elementary logic, instead of $\vdash \varphi \to \psi$ one may consider $\varphi \vdash \psi$ as well as, usually, its contrapositive $\neg\psi \vdash \neg\varphi$.[5]

The second point is more serious than the first, since this may involve other formulas than implications: the closest counterpart of *tertium non datur* $\vdash \varphi \vee \neg\varphi$ seems to be $\varphi \vdash \varphi$, which is valid in the purely partial semantics under consideration, but hardly reflects the original scheme. We will see below that there are fairly easy ways to resolve the problem of incorporating propositional logic. However, containment of tautologies may involve the restoration of the deduction theorem and so a solution to the second problem may reinforce the lurking danger observed in the third point: a revival of the omniscience connected to, for example, **K** and **I**.

The third point is a very serious one. It is also most easily overlooked, since usually we focus on principles such as **N** and **K**. To make the point entirely explicit we shortly review the deductive system which corresponds to the purely partial semantics with verificational consequence.

The core system corresponding to the purely partial semantics for modal logic consists of the rules of $\mathbf{M}^+$. We repeat some of its highlights:[6]

(R13) $B\varphi \wedge B\psi \vdash B(\varphi \wedge \psi)$ (also called $\mathbf{C_r}$)

(R15) if $\varphi \vdash \psi$ then $B\varphi \vdash B\psi$ (also called $\mathbf{I_r}$)

(R18) $B(\varphi \vee \psi) \vdash \hat{B}\varphi \vee B\psi$ (also called $\mathbf{K_r}$)

(R19) $\hat{B}(\varphi \wedge \neg\varphi) \vdash \psi$ (*modal ex falso*).

R13 is the relativized counterpart of **C**, and is therefore dubbed $\mathbf{C_r}$. An analogous similarity holds for R15, which essentially relativizes **I**. We noticed in chapter 4 that, modulo the other rules, R18 amounts to $B(\varphi \to \psi) \vdash B\varphi \to B\psi$, which is a relativized form of **K**.

Although 'omniscience' is not a disadvantageous property *per se* (its properness depending on the type of omniscience and the mode of explicit belief), we will try and see what can be achieved by extending and varying the standard semantics. Different options will be considered in the next section.

7.3 The recovery of tautologies

We know that there is a very straightforward solution to the problem how to get our cherished tautologies back, even within a purely partial semantics. As demonstrated in chapter 3 and chapter 4, we merely have to change the notion of validity from 'always true' to 'never false'. However, the problem is not really how to recover tautologies, but how to recover them without turning the logic into a normal modal system, in other words, how to avoid attributing overly strong properties to conscious belief

[5] In chapter 4 we provided these 'partial' counterparts of normal modal systems such as **S4**.

[6] See section 4.3.

and knowledge. The 'falsificational' approach indicated above in fact normalizes the logic and is therefore unfit for our purpose. A similar story can be told about the supervaluation approach[7], cf. section 3.4.6. In the rest of this section we consider two other options:

- expansion of the language with new 'classical' connectives;
- a hybrid approach to truth: simultaneous recursion on total and partial truth relations.

7.3.1 expanding the language

One of the most obvious ways to incorporate classical validities is the introduction of new propositional connectives. Now adding constants (0-place connectives) such as the primitive $\top$ increases the set of tautologies, but $\top, \neg\neg\top, \ldots, \top \vee \varphi$ hardly reveal the typical structure of classically valid formulas. We would like to mimic classical tautologies such as $\varphi \rightarrow \varphi$ by, for example, a partially valid $\varphi \supset \varphi$. But what are the truth conditions for such a non-standard implication?

In section 3.4.1 we discussed the non-standard implications $\supset$ and $\rightharpoonup$, which are defined by means of the non-standard negations $\sim$ and ∂, respectively. We noticed the classical effect of these extra connectives. Since $\rightharpoonup$ can only be defined on 4-valued models, we restrict attention to $\supset$ here. We observed that by incorporation of $\sim$ (or, equivalently, $\supset$) the classical tautologies can be regained. Moreover, in section 4.5.1 we saw that the $\neg$-free fragment of the modal language (i.e. using only $\sim, \wedge, \vee, B, \hat{B}$) is entirely classical. So there is a price to pay for capturing the tautologies in this way: if φ is valid, then $B\varphi$ is also valid. Also we have $\models B(\varphi \supset \psi) \supset (B\varphi \supset B\psi)$. In other words, **N** and **K** hold for the $\neg$-free sublanguages. Therefore, this approach is deficient for two reasons:

- we have not restored all classical tautologies, but rather have built in a set of tautologies;
- with respect to these validities, omniscience again holds, in its worst form.

We therefore continue our quest for 'safe' tautologies.

7.3.2 a hybrid approach to truth

Another way to incorporate tautologies is to adopt a dual perspective on the semantic states. Worlds as such are complete (something must be either true or false in the real world), but from the point of view of the agent they are partial: in general, she only has information about part of the world. This idea[8] can be implemented in partial

[7] At least when supervaluation depends on *internal* extensions. This and other digressions are discussed in [Th92b].

[8] Cf. section 6.3.2 for a similar total proposal due to [FH88] and section 7.5 for comparison.

semantics by distinguishing two kinds of truth relations. One is the bivalent truth relation $\models$, the other the trivalent truth relation $\mathrel{|\!\equiv}$. Their counterparts are non-truth ($\not\models$) and falsity ($\mathrel{\equiv\!|}$), respectively. A situation is necessarily true or false with respect to $\models$, but may be undefined with respect to $\mathrel{|\!\equiv}$.

The definition of a partial model $M = \langle S, \vec{\mathsf{B}}, V \rangle$ and the trivalent truth and falsity conditions (for $\mathrel{|\!\equiv}$ and $\mathrel{\equiv\!|}$) are as in the standard semantics. In addition, we have the $\models$ conditions:[9]

$$
\begin{aligned}
&s \models p \Leftrightarrow V(p, s) = 1\\
&s \models \neg\varphi \Leftrightarrow s \not\models \varphi\\
&s \models \varphi \wedge \psi \Leftrightarrow s \models \varphi \,\&\, s \models \psi\\
&s \models B_i\varphi \Leftrightarrow s \mathrel{|\!\equiv} B_i\varphi \Leftrightarrow \forall t \in \mathsf{B}_i[s] : t \mathrel{|\!\equiv} \varphi
\end{aligned}
$$

The definition of validity is also entirely straightforward:

$$\models \varphi \quad \text{iff} \quad M, s \models \varphi \text{ for all models } M \text{ and situations } s.$$

Notice that when checking the validity of a formula, we start with a two-valued evaluation and are dragged into the three-valued mode only by the modal operators. In other words, it is the doxastic operator which makes one change from objective to subjective truth, and this seems intuitively correct. Consequently, we have a partial 'internal' logic (i.e. within B) and a classical 'external' logic. Let us give some of the properties of the hybrid system. To start, notice that the relation $\models$ is indeed bivalent, and that $\mathrel{|\!\equiv}$ and $\mathrel{\equiv\!|}$ are still coherent with respect to each other, in the sense that

$$s \mathrel{|\!\equiv} \varphi \Rightarrow s \not\mathrel{\equiv\!|} \varphi$$

Coherence can be strengthened to a result that relates partial and classical truth:[10]

Proposition 7.1 (propagation) $s \mathrel{|\!\equiv} \varphi \Rightarrow s \models \varphi$.

Proof: By simultaneous induction on the structure of φ we prove that $s \mathrel{|\!\equiv} \varphi \Rightarrow s \models \varphi$ and $s \mathrel{\equiv\!|} \varphi \Rightarrow s \not\models \varphi$.

- (basic case) if $\varphi = p$ then $s \mathrel{|\!\equiv} p \Rightarrow V(p, s) = 1 \Rightarrow s \models p$ and $s \mathrel{\equiv\!|} p \Rightarrow V(p, s) = 0 \Rightarrow$ (coherence) $V(p, s) \neq 1 \Rightarrow s \not\models p$.
- (induction step) assume the claims above for certain φ and ψ (IH), then:
  - $s \mathrel{|\!\equiv} \neg\varphi \Rightarrow s \mathrel{\equiv\!|} \varphi \Rightarrow$ (IH) $s \not\models \varphi \Rightarrow s \models \neg\varphi$;
    $s \mathrel{\equiv\!|} \neg\varphi \Rightarrow s \mathrel{|\!\equiv} \varphi \Rightarrow$ (IH) $s \models \varphi \Rightarrow s \not\models \neg\varphi$.
  - $s \mathrel{|\!\equiv} \varphi \wedge \psi \Rightarrow s \mathrel{|\!\equiv} \varphi \,\&\, s \mathrel{|\!\equiv} \psi \Rightarrow$ (IH) $\Rightarrow s \models \varphi \,\&\, s \models \psi \Rightarrow s \models \varphi \wedge \psi$;
    $s \mathrel{\equiv\!|} \varphi \wedge \psi \Rightarrow s \mathrel{\equiv\!|} \varphi$ *or* $s \mathrel{\equiv\!|} \psi \Rightarrow$ (IH) $\Rightarrow s \not\models \varphi$ *or* $s \not\models \psi \Rightarrow s \not\models \varphi \wedge \psi$.

[9] There is a dual possibility for the basic clause: $s \models p \Leftrightarrow V(p, s) \neq 0$. Some reflection learns that this alternative may influence the specific evaluation, but not the class of models involved.

[10] Notice that a similar result does not hold for a 4–valued approach: coherence is of vital importance here.

– $s \models B_i\varphi \Leftrightarrow s \models B_i\varphi$;
$s \dashv B_i\varphi \Rightarrow \exists t \in \mathsf{B}[s] : t \dashv \varphi \Rightarrow$ (IH) $\exists t \in \mathsf{B}[s] : t \not\models \varphi \Rightarrow s \not\models B_i\varphi$. ■

Similar to the purely partial semantics we do not obtain internal persistence. Fortunately, the revealing connection between awareness and explicit belief known from the special awareness logic is conserved: if someone is aware of a tautology, he believes it.

Proposition 7.2
If φ is purely propositional and $\models \varphi$, then $\models A\varphi \rightarrow B\varphi$.

Proof: (by contraposition) Let φ be a propositional formula such that $\not\models A\varphi \rightarrow B\varphi$. Then there is a model $M = \langle S, \mathsf{B}, V\rangle$ and a state $s \in S$ such that $V(p_i, t) \neq \frac{1}{2}$ for all $t \in \mathsf{B}[s]$ and all atoms $p_1, \ldots, p_n$ occurring in φ, and, moreover, that $M, t' \not\models \varphi$ for some $t' \in \mathsf{B}[s]$. By induction it follows that for each $\psi \in \mathcal{L}^0\{p_1, \ldots, p_n\}$ and $t \in \mathsf{B}[s]$: $M, t \models \psi$ or $M, t \dashv \psi$. Consequently $M, t' \dashv \varphi$, and therefore by propagation $M, t' \not\models \varphi$, thus $\not\models \varphi$. ■

The 'core logic' of the hybrid semantics consists of classical propositional logic, the conjunction scheme **C** and the rule **I** restricted to modal strong consequence (with respect to $\mathbf{M}^+$):

$\mathbf{I}_{\mathbf{M}^+}$ if $\varphi \vdash_{\mathbf{M}^+} \psi$ then $\vdash B\varphi \rightarrow B\psi$

C $\vdash (B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$

Further properties of explicit belief are triggered by suitable conditions on the frame. In general, the framework of hybrid truth is a rather flexible one, more or less comparable to standard possible world semantics. A remarkable exception to this is the **5** scheme of *negative introspection*. The corresponding condition is extremely strong: accessibility has to be both *euclidean* and lead to total situations, i.e. states such that every formula is either supported or rejected. This heavy constraint turns the logic into a normal modal system (**NK5**), which, as we have seen, is unfit for our enterprise. Perhaps the right conclusion from this is that requiring negative introspection for explicit belief is very nonsensical, and we should be punished for such a sin. But then we may argue that *positive introspection* is almost equally counterintuitive for explicit belief.[11]

We may therefore conclude that the hybrid semantics is partly successful. A number of problems is solved more or less automatically. In particular, we do have tautologies, but we do not have **N**-omniscience. However, we have to confess that some of the properties attributed to belief in this way are less fortunate. One of the chief points is the remaining **K**–omniscience, which is related to the core logic of the hybrid semantics: in the hybrid system people are forced to believe the conclusions derivable within their own belief. Unlike other authors we do not think that such closure under implication is acceptable for *any* sense of explicit belief. We will turn to this problem in the next section.

[11] Perhaps the converse schemata: $\vdash BB\varphi \rightarrow B\varphi$ ($\mathbf{4_c}$) and $\vdash B\neg B\varphi \rightarrow \neg B\varphi$ ($\mathbf{5_c}$), called 'extraspection' schemata in [vdH91], are preferable to the usual forms of introspection.

7.4 The elimination of residual omniscience

Despite the relative success of modifications of partial semantics such as the hybrid approach, we are left with some forms of omniscience. This is sometimes argued to be inevitable: if the logic is to contain more than just the modal substitution instances of the classical propositional calculus, then these extra principles would lead to new belief or knowledge, i.e. create omniscience. This argument is not conclusive, however. The point is that the derived belief may be intuitively acceptable, and this is what determines whether the type of inference is acceptable for explicit belief; if not, we arrive at a form of omniscience that should be exorcized. To wit,

C$_c$ $\quad \vdash B(\varphi \wedge \psi) \rightarrow (B\varphi \wedge B\psi)$

seems fully acceptable for explicit belief or human knowledge.[12]

From the preceding sections it emerges that the principles underlying residual omniscience may be of one of the following shapes:

- the inference rule **I**, possibly restricted to strong consequences (i.e. 'tautological entailments', *ex falso* and their modal counterparts);
- the schemata **C** and **K**, which both combine different beliefs into one.

Recall from previous chapters that

I $\quad \vdash \varphi \rightarrow \psi \Rightarrow \vdash B\varphi \rightarrow B\psi,$

C $\quad \vdash (B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$

K $\quad \vdash B(\varphi \rightarrow \psi) \rightarrow (B\varphi \rightarrow B\psi)$

Before we start our inquiry into an adequate partial semantics for logics avoiding these principles, we make some general observations.

First, notice that, *modulo* the accepted principle **C**$_c$, **I** is equivalent to the 'extensionality principle' **E**.

E $\quad \vdash \varphi \leftrightarrow \psi \Rightarrow \vdash B\varphi \leftrightarrow B\psi$

This equivalence even holds for the relativized counterparts of these principles, with **rL** as propositional background logic:

Proposition 7.3
EC$_c \overset{\mathbf{pL}}{\Leftrightarrow}$ **I**, *and similarly for relative counterparts:* **EC**$_c \overset{\mathbf{rL}}{\Leftrightarrow}$ **I**.

Proof: the **pL**-equivalence is from [Ch80, theorem 8.11(1)], the **rL**-equivalence is similar. ■

Second, the combination schemata are also intimately related:

[12] Yet **C**$_c$ is rejected for resource bounded knowledge by [Mo88] in a computational setting.

Proposition 7.4 **K** *and* **C** *are equivalent modulo* **pL** *and* **I**.

Proof: from [Ch80], theorem 8.11(2)+ exercise 8.13; also compare the proof of proposition 4.1. ∎

Notice that the latter equivalence only holds in the presence of **I**; in its absence, **C** and **K** have to be considered separately.

The principles **I**, **K** and **C** will be studied in the rest of this section. We will confront these principles with two ideas as to how to deal with them:

- recycling awareness
- superimposing awareness

We wil relate these ideas to the different strategies discussed earlier, although we restrict ourselves to the standard connectives.

7.4.1 recycling awareness

The first proposal is especially directed to the elimination of rule **I** for explicit belief[13] (or $\mathbf{I_r}$ restricted to $\mathbf{M}^+$ in the hybrid case). It is easy to devise a semantics to this effect: simply block the truth conditions for modal operators, i.e. regard formulas of the form $B\varphi$ as atomic. This approach is not very flexible, however. And before we reject the entire inference rule, notice that there are instances of **I** which are fully acceptable: $\mathbf{C_c}$ is one of them! Also, **I** applied to *ex falso* produces

$$\vdash B(\varphi \wedge \neg\varphi) \rightarrow B\psi$$

which again is acceptable, since this is a direct propositional consequence of the consistency of belief:

D* $\quad \vdash \neg B(\varphi \wedge \neg\varphi)$

Nobody explicitly believes contradictions — one's private beliefs may be mutually inconsistent, but that is a different matter.

Applied to other propositional laws, such as *double negation*, the acceptability becomes less evident. So,

$$\vdash B\neg\neg\varphi \rightarrow B\varphi$$

seems acceptable for some senses of belief, but not for all. Still this is an innocent case compared with truly dubious instances such as

$$\vdash B\varphi \rightarrow B(\psi \vee \neg\psi)$$

[13]The approach suggested in this subsection is in fact independent of the nature of the semantics, and may therefore also be applied to e.g. special awareness logic.

and

$$\vdash B\varphi \rightarrow B(\varphi \vee \psi).$$

The former is excluded in, for example, the hybrid system, since there **I** is restricted to strong consequences, and *tertium non datur* is not one of these.

The latter is not excluded by the hybrid system. In [LL88], Levesque & Lakemeyer propose a solution to this problem. It is based on the fact that awareness of φ need not involve awareness of $\varphi \vee \psi$. This also holds for the notion of derived awareness defined earlier. We repeat its definition here.

$$A\varphi = \bigwedge_{p \text{ in } \varphi} B(p \vee \neg p)$$

So, let us define a new notion of explicit belief B_i^A, henceforth loosely referred to as 'actual belief'

$$B_i^A\varphi = B_i\varphi \wedge A_i\varphi$$

With respect to actual belief, some of the worst results due to **I** are eliminated. In general, we have that disjunctive weakening does not hold for B^A, for example, $B^A p \rightarrow B^A(p \vee q)$ is not valid. Yet, there are almost equally dubious results which are still validated by the augmented system, e.g. $\vdash B^A p \rightarrow B^A(p \vee \neg p)$.

Moreover, though the redefined notion of explicit belief may help for a number of problematic cases related to **I**, it is easily verified that the other problematic principles, **C** and **K**, still exist.

The next subsection deals with a more radical strategy: use the awareness sieves introduced in section 6.5.

7.4.2 superimposing awareness sieves

In the previous section we found that 'recycling derived awareness' does not provide a satisfactory solution for all kinds of residual omniscience, though the approach may be acceptable for a special sense of awareness (acquaintance). We would therefore like to combine a general semantics based on partial valuations with a mechanism that controls the notion of *actual belief*.

Within the area of possible world semantics, such a flexible framework was outlined in chapter 6. It was essentially a generalization of Fagin & Halpern's logic of 'general awareness', without their structural conditions *seriality*, *transitivity*, and *euclidicity*. This so-called *sieve semantics* turned out to be a very general and flexible framework for weak modal logics. The idea is now to superimpose the awareness sieve on the hybrid partial semantics. Of course, 'recycled' awareness can also be construed as an awareness filter, but its special structure does not warrant sufficient flexibility.

A partial sieve model $M = \langle S, \vec{\mathsf{B}}, \vec{\mathcal{A}}, V\rangle$ with hybrid evaluation is defined as follows. The trivalent truth/falsity relations ($\models$ and $\Dashv$) and the bivalent truth relation ($\models$) are defined as in the hybrid semantics from section 7.3.2 (for $\mathcal{L}_{\vec{B}}$). In addition we have clauses for the *actual belief* operators B_i^A, which are now interpreted by means of awareness sieves and accessibility relations. For each i and s the awareness sieve is a set of formulas, i.e. $\mathcal{A}_i(s) \subseteq \mathcal{L}$. Here we consider the logical language $\mathcal{L}$ to be built up in the usual way from propositional variables, connectives and the modal operators B_i and B_i^A. The evaluation procedure for B_i is as before, and the additional clauses for B_i^A are:

$$s \models B_i^A\varphi \Leftrightarrow s \models\!\!\models B_i^A\varphi \Leftrightarrow s \models\!\!\models B_i\varphi \;\&\; \varphi \in \mathcal{A}_i(s)$$
$$s \rightleftharpoons\!\!| B_i^A\varphi \Leftrightarrow s \rightleftharpoons\!\!| B_i\varphi \text{ or } \varphi \notin \mathcal{A}_i(s)$$

Validity is still defined as universal bivalent truth. Here are a number of observations which indicate that the 'hybrid sieve' semantics fulfils the requirements of a proper partial interpretation:

- as before, we have coherence: $s \models\!\!\models \varphi \Rightarrow s \not\rightleftharpoons\!\!| \varphi$ for every φ and s;
- moreover, there is propagation: $s \models\!\!\models \varphi \Rightarrow s \models \varphi$;
- another useful property, also exhibited by the previous partial logics, is what may be called *classical closure*[14]: if V is bivalent for *all* situations, then $s \models\!\!\models \varphi \Leftrightarrow s \models \varphi$;
- the semantics is still *externally persistent*: extension of the valuation for a fixed frame (to which the awareness sieve belongs) implies preservation of trivalent truth and falsity.

Is this semantics as general and flexible as total sieve semantics? In other words, can we still capture every logic which extends classical propositional logic? This question is answered in the affirmative.

Theorem 7.1
Hybrid sieve semantics is sound and complete for every modal system extending **pL** *by at least the axiom scheme* $\vdash B_i^A\varphi \rightarrow B_i\varphi$.

Proof: Let S be a logic that extends **pL**. So, S is a set of formulas of the language $\mathcal{L}_{\vec{B},\vec{B}^A}$ that contains all the substitution cases of tautologies. S is characterized by the total canonical model $\mathcal{M} = \langle \mathcal{W}, \vec{\mathcal{R}}, \vec{\mathcal{A}}, \mathcal{V}\rangle$ described in the proof of corollary 6.2. By the observed property of classical closure, the total canonical model belongs to the hybrid sieve semantics. This guarantees soundness and completeness by the usual truth lemma and Lindenbaum lemma. ∎

So, every modal logic that contains all tautologies can be captured by a suitable class of models. This notion of 'modal logic' is very wide: for example the principle of extensionality (**E**) need not hold. Also, the notion of completeness is not very restricted; in its generality the previous theorem is almost void. As in the area normal modal logic, we are in general more interested in what may be called frame

[14] Cf. [vB88], and chapter 1 here.

completeness. Since we know there are normal modal systems which are incomplete with respect to the class of verifying frames[15], a similar result may hold for the present weak systems and semantics. If the sieve counts as part of the frame, it is possible to single out a number of corresponding conditions for some intuitively valid principles, such as **D*** and $\mathbf{C_c}$ for B_i^A.

- $\mathbf{C_c}$ for B_i^A, i.e. $\models B_i^A(\varphi \wedge \psi) \rightarrow (B_i^A\varphi \wedge B_i^A\psi)$ is captured by the condition that $\varphi \wedge \psi \in \mathcal{A}_i(s) \Rightarrow \varphi, \psi \in \mathcal{A}_i(s)$.
- **D*** for B_i^A, i.e. $\models \neg B_i^A(\varphi \wedge \neg\varphi)$ is captured by the condition that $\varphi \wedge \neg\varphi \notin \mathcal{A}_i(s)$.
- **D** (or, equivalently, **D***) for B_i, i.e. $\models \neg B_i(\varphi \wedge \neg\varphi)$ is captured by the condition that B_i is serial: $\forall s \exists t : s\mathsf{B}_i t$.
- **C** and $\mathbf{C_c}$ hold automatically for B_i.

These axioms show an interesting interplay. For example, the different consistency axioms for B_i^A are related to the one for B_i: if B_i is serial then **D** and **D*** are valid, and so is the propositional consequence $\neg(B_i\varphi \wedge \neg B_i\varphi)$. But then, by $B^A \Rightarrow B$, we obtain $\neg(B_i^A\varphi \wedge \neg B_i^A\varphi)$, i.e. **D** holds for B^A as well. By $\mathbf{C_c}$ this also implies **D***: $\neg B_i^A(\varphi \wedge \neg\varphi)$. The moral from this is that if we want but a few principles to be valid for B^A, the logic for B also has to be weakened, which may be beyond expectation. However, although usually *less* explicit belief is connected to a *more* idealized (i.e. stronger) logic, we do not see any *a priori* reason that this has to be the case. For example, one may implicitly believe a contradiction, without being aware of it. But when it comes to explicitly believing something, for example by expressing that belief, we should not allow inconsistencies (cf. chapter 5). Therefore, it seems preferable to require **D*** for B^A, but not for B.

7.5 Partial and total approaches compared

7.5.1 general awareness vs hybrid sieve system

It follows from corollary 6.2 and theorem 7.1 that total sieve semantics and hybrid partial semantics with a superimposed awareness sieve are extensionally equivalent, in the sense that the two approaches model the same logics for the restricted language $\mathcal{L}_{\vec{B^A}}$. Despite this technical equivalence there are differences in underlying intuitions, especially with regard to the way in which intuitively unacceptable principles are circumvented: part of the awareness which deals with knowledge of the objects and notions involved, i.e. with the conceptual information present in the agent, is accounted for by means of partiality. Another type of awareness, corresponding to what the agent actually thinks of at a certain moment, is accounted for by means of the awareness

[15]See Fine(1974), S.K. Thomason (1974), van Benthem (1978,1979,1984) for ever simpler frame-incomplete systems.

sieve. The awareness sieve is only effective on propositions which are (partially) true, and this accords the intuition that we need basic conceptual knowledge before actually being aware of something.

Of course there are other differences as well. The original logic of general awareness in [FH88] contains the operators L_i, A_i and B_i. In what we loosely called 'the hybrid sieve semantics' there are the operators B_i and B_i^A. Moreover, the B–operator of the former approach corresponds to the B^A–operator of the latter. Finally, the R_i relation in models of general awareness interprets L_i, whereas the hybrid sieve relation B_i deals with B_i. So there is a clear gap between the two specific systems. Can we bridge the gap?

First, addition of the operators L_i for implicit belief to the hybrid sieve system is quite straightforward. The model-theoretic counterpart of L_i is an accessibility relation L_i (corresponding to R_i) such that $\mathsf{L}_i \subseteq \mathsf{B}_i$. If we want our hybrid sieve logic to coincide with the logic of general awareness, L_i has to be subjected to the same conditions as R_i: it should be serial, transitive and euclidean. The new evaluation conditions for L_i are:

$$s \models L_i\varphi \Leftrightarrow \forall t \in \mathsf{L}_i[s] : t \models \varphi$$
$$s \rightrightarrows\!\!| \; L_i\varphi \Leftrightarrow \exists t \in \mathsf{L}_i[s] : t \rightrightarrows\!\!| \; \varphi$$
$$s \vDash L_i\varphi \Leftrightarrow \forall t \in \mathsf{L}_i[s] : t \vDash \varphi$$

It should be clear from the $\vDash$ clause and validity concept of the hybrid approach that the logic for L_i will be entirely normal, and due to the accessibility constraints on L_i, its modal system will be (N)**KD45**.

Second, the awareness operators A_i could also be added, with the following simple truth/falsity conditions:

$$s \models A_i\varphi \Leftrightarrow s \vDash A_i\varphi \Leftrightarrow \varphi \in \mathcal{A}_i(s)$$
$$s \rightrightarrows\!\!| \; A_i\varphi \Leftrightarrow \varphi \notin \mathcal{A}_i(s)$$

Then B^A could be redefined by $B_i^A\varphi = B_i\varphi \wedge A_i\varphi$. We did not take this road in the previous section, since

- we already have a (derived) awareness operator A_i, defined by $A_i\varphi = \bigwedge\!\!\!\bigwedge B_i(p \vee \neg p)$ over all p in φ;
- addition of new awareness operators A_i would lead to unacceptable interaction with the B_i–operators: $B_i(A_i\varphi \vee \neg A_i\varphi)$ would be validated, which seems intuitively wrong. (Notice this problem does not occur for B^A: $\not\vDash B_i^A(A_i\varphi \vee \neg A_i\varphi)$.) It is technically possible to avoid bivalence of A_i by partializing the awareness sieves, i.e. duplicate the sieve function $\mathcal{A}_i$ into the pair $\mathcal{A}_i^+, \mathcal{A}_i^-$, where $\mathcal{A}_i^+(s) \subseteq \mathcal{L}$ and $\mathcal{A}_i^-(s) \subseteq \mathcal{L}$ and give appropriately modified truth/falsity conditions for the awareness operator. We have no intuitions about 'negative awareness' different from lack of awareness, however.

Third, if the actual belief operator B^A has to correspond to B in the logic of general awareness, the operator B from hybrid sieve logic has to have a counterpart in general awareness logic. It is possible to add such intermediating operators B_i^l for each i, and duplicate the awareness sieves such that B_i^l is interpreted by means of $\mathcal{A}_i^l$. A suitable transformation of a hybrid sieve model into a GAL model is also feasible.

7.5.2 special awareness logic and partial systems

Without the awareness sieves we can also compare total and partial awareness logics.[16] In many respects SAL is similar to the hybrid system. Both approaches are characterized by a twofold perspective on truth: total and partial. Are the hybrid system and SAL equivalent?

Indeed the two approaches share a large number of properties. For example, both have the rule $\mathbf{I}_{\mathbf{M}^+}$, the axiom scheme **C**, **D** is modelled by seriality of R_i and B_i, and for serial models a strong possibility rule applies:

$\mathbf{P}^* \models \varphi \Rightarrow \models \hat{B}_i B^* \varphi,$

where B^* abbreviates a sequence of operators from $\{B_1, \ldots, B_m\}$.[17] We also have proposition 7.2 for both systems. Although we have to impose, apart from transitivity, the additional condition of upward monotonicity on the models to capture positive introspection (**4**), both systems 'crash' when requiring negative introspection (**5**).

All these properties refer to *explicit* belief; we can add operators for *implicit* belief in the manner of the previous discussion on total and hybrid sieve semantics.

Given this similarity, it may not be surprising that every hybrid validity is also provably an SAL validity. First we describe a transformation and prove it preserves truth. Let $M = \langle W, \vec{\mathsf{R}}, \vec{\mathcal{A}}, V\rangle$ be an SAL model (see section 6.3.2.) We can construct a corresponding hybrid sieve model $M' = \langle S, \vec{\mathsf{B}}, \vec{\mathsf{L}}, V'\rangle$ [18] with: (Ψ is an arbitrary subset of $\mathcal{L}_{\neg,\wedge,\vec{B},\vec{L}}(Prop)$)

- $S = W \times \wp(Prop)$ ($\langle w, \Psi\rangle$ will be written as w_Ψ)
- $w_\Psi \mathsf{B}_i v_{\Psi\cap\mathcal{A}_i(w)} \Leftrightarrow w\mathsf{R}_i v$
- $w_\Psi \mathsf{L}_i v_\Psi \Leftrightarrow w\mathsf{R}_i v$
- $V'(p, w_\Psi) = \begin{cases} V(p,w) & \text{iff } p \in \Psi \\ \frac{1}{2} & \text{otherwise} \end{cases}$

---

[16] In fact an initial goal of SAL was to simulate Levesque's [Le84a] logic in augmented possible worlds semantics.

[17] By iterative use of $\mathbf{P}^*$ one can derive a generalization of $\mathbf{P}^*$ in which the dual operators $\hat{B}_1, \ldots, \hat{B}_m$ are also allowed in the sequence B^*.

[18] For the language restricted to explicit belief $\langle W, \vec{\mathsf{B}}, V'\rangle$ suffices.

Lemma 7.1

$$\begin{array}{ll} (i) & M,w \models^{\Psi} \varphi \Leftrightarrow M',w_{\Psi} \models\!\!\!\equiv \varphi \\ (ii) & M,w =\!\!\!| ^{\Psi}\varphi \Leftrightarrow M',w_{\Psi} \equiv\!\!\!| \varphi \\ (iii) & M,w \models \varphi \Leftrightarrow M',w_{Prop} \models \varphi \end{array}$$

Proof: by induction on the structure of φ. First consider $p \in Prop$:

p: (i) $M,w \models^{\Psi} p \Leftrightarrow V(p,w) = 1 \,\&\, p \in \Psi \Leftrightarrow V'(p,w_{\Psi}) = 1 \Leftrightarrow M',w_{\Psi} \models\!\!\!\equiv p$;
(ii) $M,w =\!\!\!| ^{\Psi} p \Leftrightarrow V(p,w) = 0 \,\&\, p \in \Psi \Leftrightarrow V'(p,w_{\Psi}) = 0 \Leftrightarrow M',w_{\Psi} \equiv\!\!\!| p$;
(iii) $M,w \models p \Leftrightarrow V(p,w) = 1 \Leftrightarrow V'(p,w_{Prop}) = 1 \Leftrightarrow M',w_{Prop} \models p$;

Now assume the induction hypothesis (IH) for φ and ψ, then:

$\neg$: (i) $M,w \models^{\Psi} \neg\varphi \Leftrightarrow M,w =\!\!\!| ^{\Psi}\varphi \overset{\text{IH}}{\Leftrightarrow} M',w_{\Psi} \equiv\!\!\!| \varphi \Leftrightarrow M',w_{\Psi} \models\!\!\!\equiv \neg\varphi$;
(ii) $M,w =\!\!\!| ^{\Psi}\neg\varphi \Leftrightarrow M,w \models^{\Psi} \varphi \overset{\text{IH}}{\Leftrightarrow} M',w_{\Psi} \models\!\!\!\equiv \varphi \Leftrightarrow M',w_{\Psi} \equiv\!\!\!| \neg\varphi$;
(iii) $M,w \models \neg\varphi \Leftrightarrow M,w \not\models \varphi \overset{\text{IH}}{\Leftrightarrow} M',w_{Prop} \not\models\!\!\!\equiv \varphi \Leftrightarrow M',w_{Prop} \models \neg\varphi$;

$\wedge$: (i) $M,w \models^{\Psi} \varphi\wedge\psi \Leftrightarrow M,w \models^{\Psi} \varphi \,\&\, M,w \models^{\Psi} \psi \overset{\text{IH}}{\Leftrightarrow} M',w_{\Psi} \models\!\!\!\equiv \varphi \,\&\, M',w_{\Psi} \models\!\!\!\equiv \psi \Leftrightarrow$ $M',w_{\Psi} \models\!\!\!\equiv \varphi \wedge \psi$;
(ii) $M,w =\!\!\!| ^{\Psi}\varphi\wedge\psi \Leftrightarrow M,w =\!\!\!| ^{\Psi}\varphi$ *or* $M,w =\!\!\!| ^{\Psi}\psi \overset{\text{IH}}{\Leftrightarrow} M',w_{\Psi} \equiv\!\!\!| \varphi$ *or* $M',w_{\Psi} \equiv\!\!\!| \psi \Leftrightarrow$ $M',w_{\Psi} \equiv\!\!\!| \varphi \wedge \psi$;
(iii) $M,w \models \varphi \wedge \psi \Leftrightarrow M,w \models \varphi \,\&\, M,w \models \psi \overset{\text{IH}}{\Leftrightarrow}$
$M',w_{Prop} \models \varphi \,\&\, M',w_{Prop} \models \psi \Leftrightarrow M',w_{Prop} \models \varphi \wedge \psi$;

B: (i) $M,w \models^{\Psi} B_i\varphi \Leftrightarrow \forall v \in \mathsf{R}_i[w] : M,v \models^{\Psi\cap\mathcal{A}_i(w)} \varphi \Leftrightarrow$ (IH+def.B)
$\forall v_{\Psi\cap\mathcal{A}_i(w)} \in \mathsf{B}_i[w_{\Psi}] : M',v_{\Psi\cap\mathcal{A}_i(w)} \models\!\!\!\equiv \varphi \Leftrightarrow M',w_{\Psi} \models\!\!\!\equiv B_i\varphi$;
(ii) $M,w =\!\!\!| ^{\Psi}B_i\varphi \Leftrightarrow \exists v \in \mathsf{R}_i[w] : M,v =\!\!\!| ^{\Psi\cap\mathcal{A}_i(w)}\varphi \Leftrightarrow$ (IH+def.B)
$\exists v_{\Psi\cap\mathcal{A}_i(w)} \in \mathsf{B}_i[w_{\Psi}] : M',v_{\Psi\cap\mathcal{A}_i(w)} \equiv\!\!\!| \varphi \Leftrightarrow M',w_{\Psi} \equiv\!\!\!| B_i\varphi$;
(iii) $M,w \models B_i\varphi \Leftrightarrow \forall v \in \mathsf{R}_i[w] : M,v \models^{\mathcal{A}_i(w)} \varphi \Leftrightarrow$ (IH+def.B)
$\forall v_{\mathcal{A}_i(w)} \in \mathsf{B}_i[w_{Prop}] : M',v_{\mathcal{A}_i(w)} \models\!\!\!\equiv \varphi \Leftrightarrow M',w_{Prop} \models\!\!\!\equiv B_i\varphi \Leftrightarrow M',w_{Prop} \models B_i\varphi$;

L: (i) $M,w \models^{\Psi} L_i\varphi \Leftrightarrow \forall v \in \mathsf{R}_i[w] : M,v \models^{\Psi} \varphi \Leftrightarrow$ (IH+def.L)
$\forall v_{\Psi} \in \mathsf{L}_i[w_{\Psi}] : M',v_{\Psi} \models\!\!\!\equiv \varphi \Leftrightarrow M',w_{\Psi} \models\!\!\!\equiv L_i\varphi$;
(ii) $M,w =\!\!\!| ^{\Psi}L_i\varphi \Leftrightarrow \exists v \in \mathsf{R}_i[w] : M,v =\!\!\!| ^{\Psi}\varphi \Leftrightarrow$ (IH+def.L)
$\exists v_{\Psi} \in \mathsf{L}_i[w_{\Psi}] : M',v_{\Psi} \equiv\!\!\!| \varphi \Leftrightarrow M',w_{\Psi} \equiv\!\!\!| L_i\varphi$;
(iii) $M,w \models L_i\varphi \Leftrightarrow \forall v \in \mathsf{R}_i[w] : M,v \models \varphi \Leftrightarrow$ (IH+def.L)
$\forall v_{Prop} \in \mathsf{L}_i[w_{Prop}] : M',v_{Prop} \models \varphi \Leftrightarrow M',w_{Prop} \models L_i\varphi$;

■

Notice that the lemma does not claim full equivalence of the models involved; in fact this will be hard to achieve, since $M',w_{\Psi} \models \varphi$ has no clear counterpart in M if $\Psi \subset Prop$. The lemma shows containment of the set of SAL validities in the set of the hybrid validities:

Theorem 7.2 *Every formula valid with respect to the hybrid partial semantics is also valid with respect to* SAL *semantics.*

Proof: by contraposition, if φ is not valid in SAL semantics, then for some model M and world w: $M,w \not\models \varphi$. So, by the lemma above, there is a hybrid model M' such that $M',w_{Prop} \not\models \varphi$, and therefore φ is invalid for the hybrid semantics. ■

Does the converse of this theorem hold as well? No, it does not! The main reason is that the hybrid semantics is more permissive. Unlike the SAL models it does not transfer the set of defined propositional atoms from a situation to its doxastic alternatives. The effect of this difference can be seen in a formula such as:

$$B_i B_j(p \vee \neg p) \rightarrow B_i(p \vee \neg p).$$

This formula is valid in SAL; by invoking the notion of derived awareness, we even have $\models B_i A_j \varphi \rightarrow A_i \varphi$ in SAL.[19] These formulas are invalid in the hybrid system. It is not entirely clear to us whether we should desire the validity of the displayed formula — it may depend on the notion of awareness involved.

In all, despite these minor differences SAL and the hybrid semantics are very similar. We do believe the hybrid models to be more natural, however, since there is no need to specify more of the content of an alternative doxastic state than the agent is aware of. Although we are not claiming 'psychological reality' for any of the proposals made here, it is clear, we think, which approach is more intuitive in this respect.

hybrid semantics vs expanded language

There is one other connection to hybrid semantics that deserves attention: its relation to the expanded language.[20] The idea is that the external bivalence can also be triggered by giving the right scope to non-standard connectives, say $\sim$. So, let us try to give a translation $'$ from $\varphi \in \mathcal{L}_{\neg,\wedge,\vec{B},\vec{L}}$ into $\varphi' \in \mathcal{L}_{\neg,\sim,\wedge,\vec{B},\vec{L}}$ such that $\models \varphi \Leftrightarrow \models\!\!\equiv \varphi'$, where the former validity is in hybrid semantics, and the latter in the standard semantics.

$$p' = p \quad (\neg\varphi)' = \sim\varphi' \quad (\varphi \wedge \psi)' = (\varphi' \wedge \psi') \quad (B\varphi)' = B\varphi \quad (L\varphi)' = L\varphi'$$

The simple effect of this is that each $\neg$ occurring outside the scope of all B's is changed into $\sim$, and each occurrence within B is left as it is. This already provides a simple but insightful observation:

Proposition 7.5 *For each $\varphi \in \mathcal{L}_{\neg,\wedge,\vec{B},\vec{L}}$: $\models \varphi \Leftrightarrow \models\!\!\equiv \varphi'$*

Proof: let M be a partial Kripke model. Then $M, s \models \varphi \Leftrightarrow M, s \models\!\!\equiv \varphi'$ for all situations s in M, which can be shown by induction on the complexity of φ ■

Since a reverse embedding does not seem obtainable, this shows on the one hand that the hybrid approach is somewhat more restricted than the method of expanding the language. On the other hand, the proposition also indicates that the expanded language contains interesting fragments, such as the one provided by the translation, with $\sim$outside of B_i and $\neg$ inside of it.

---

[19] For suppose $w \models B_i A_j \varphi$, then for arbitrary p in φ and $v \in \mathsf{R}_i[w]$: $v \models^{\mathcal{A}_i(w)} B_j(p \vee \neg p) \Rightarrow u \models^{\mathcal{A}_i(w) \cap \mathcal{A}_j(v)} p \vee \neg p$ for all $u \in \mathsf{R}_j[v] \Rightarrow p \in \mathcal{A}_i(w) \cap \mathcal{A}_j(v)$ and therefore $p \in \mathcal{A}_i(w)$, thus $v \models^{\mathcal{A}_i(w)} p \vee \neg p \Rightarrow w \models A_i \varphi$.

[20] Following a suggestion by van Benthem.

7.6 Conclusion

In this chapter we first reconsidered standard partial semantics as a candidate for an adequate logical description of awareness. Indeed this already provides a rather weak logic, as is required for conscious belief. So, literally speaking, a number of problematic principles, which are loosely called logical omniscience, are circumvented. However, the standard partial approach suffers from two major problems: one is that intuitively valid forms such as $Bp \vee \neg Bp$ are also eliminated, the other is that most types of logical omniscience come back in relativized form, e.g. $Bp \Rightarrow B(p \vee q)$. This may be paraphrased by saying that in the standard approach the external part of the awareness logic (which we wish to be classical) is too weak, and the internal part (which we wish to be non-classical) is still too strong.

We therefore proceeded by trying to eliminate these drawbacks. Starting with the first major problem, one possible solution was to reinsert into the standard language some of the non-standard connectives introduced in chapter 1. We noticed that, though there are classical sublanguages within the expanded language, this does not remove invalid formulas as $Bp \vee \neg Bp$. One may suggest at this point that this objection can be countered by giving 'believe or not believe p' a different translation, e.g. $Bp \vee {\sim}Bp$. Then we considered another strategy: similar to the (special) awareness logic discussed in chapter 6, have different truth relations within one and the same semantics. We argued that such a hybrid approach successfully deals with the problem of the 'missing tautologies'. Later it turned out that this hybrid logic can be embedded into the logic of the expanded language. So there is a fragment in the expanded language that corresponds to the standard formulas in the sense that classical truth of a formula in the hybrid system is equivalent to partial truth of its 'expanded' translation.

None of these proposals, however, solves the second major problem. Again we considered two ways of dealing with the problem of 'residual omniscience'. One possibility was to derive awareness from explicit belief, and then add this awareness to explicit belief. Though a nice and concise option, the effect of this turned out to be limited. So, we moved to a more demanding attack: add the syntactic awareness sieves to the hybrid system. We showed that again the resulting semantics is fully flexible in the sense that every modal logic extending the classical propositional calculus can be modelled. Again, for specific applications, we may constrain the admissible awareness sieves by conditions reminiscent of neighbourhood semantics.

Finally, a number of connections between partial and total approaches to awareness were made. In particular we showed that the semantics of the special awareness logic can be embedded in the hybrid semantics proper. We claimed that by virtue of built-in partiality, hybrid sieve semantics is more natural than total sieve semantics, yet the greater freedom allowed in hybrid sieve models might be considered problematic.

In fact, already for sieve semantics there is considerable liberty in the way validities can be described.[21] Now at first sight the situation is even worse for hybrid sieve

[21] For example, the modal scheme **4** can be modelled *both* by the class of Kripke frames with empty accessibility combined with an awareness sieve filtering out formulas of the form $B\varphi \rightarrow BB\varphi$, *and* by the class of transitive Kripke frames with the coarsest sieve, that lets all formulas pass.

semantics: partiality may also account for non-valid principles. Yet, we argue that this extra possibility is not disadvantageous, since the non-validity that is obtained by partiality is 'for free', as it were. So there is a clear order that determines the locus of explanation or description of (in)valid principles: first partiality (which is always present), then accessibility (constrained by general conditions) and finally, as a last escape route, by a stipulation on the awareness sieve. Therefore, hybrid sieve models are at least not inferior to total sieve models in this respect.

Since the class of hybrid sieve models is larger than the class of total sieve models, this also seems to present a computational disadvantage. Though this may sound paradoxically, this need not be true, as we will see in the next part of this thesis.

Part III

Models for Computer Knowledge

Introduction to part III

In the last part of this thesis we will try and see to what extent (partial) modal logic can be used to store and retrieve knowledge in computer systems. By storing enough knowledge into a computer system, we can, in principle, make it easier for the ordinary user to retrieve information from such a system. The idea is not merely to have enough basic data in the system, but to let the computer reason about its own knowledge[22], as if it were a human agent. So, we want the system to introspect its knowledge. Moreover, the system has to know about the knowledge and the ignorance of the user, in order to provide pragmatically sound answers.

For example, suppose some tour office wonders if a flight information system knows whether there are regular flights from Amsterdam to Bejing. Now 'Do you know whether there are flights from Amsterdam to Bejing' can properly be answered by either 'Yes' or perhaps 'Yes, every Tuesday and Friday', whereas the same question posed by a tourist longing for a vacation cannot be answered by 'Yes', since this would not be very cooperative, and after all, that is what the system is for. Instead, 'Yes, there is one tomorrow. Do you want to make a reservation?' would be a pragmatically sound continuation of the dialogue in this case. So the point is that the system has to know both about the data and its knowledge thereof, and about the knowledge of the user and the things she would like to know. Given these desiderata, the performance of present day information systems can hardly be called satisfactory. This is partly caused by the problem how to store all relevant knowledge of the real world into the computer (for which facts *are* relevant?), and partly by the difficulty of knowing the user's state of mind, and her knowledge and ignorance. The general idea is that epistemic logic may be helpful to describe world knowledge, as well as knowledge about knowledge and ignorance, etcetera.

There are, essentially, two different ways of using epistemic logic to store knowledge. One is the *syntactic* approach, which is mostly adopted. In this method the domain knowledge is represented by a set of formulas, a query (or, rather, its propositional contents) is translated into a logical formula, and the program tries to deduce this formula (or its negation) from the basic knowledge. Such theorem proving however suffers from problems of undecidability (of first-order logic) or intractibility (of certain modal logics).

Therefore the other way of storing information is studied here. In this method, that may loosely be called the *semantic* approach, the knowledge is stored in a (logical) model. Then answering a query amounts to merely checking whether or not the formula representing the query is true. This model checking is usually relatively easy, since evaluation of a modal formula at a possible world is determined by the model structure and the recursive truth conditions. Following the structure of the formula the truth or falsity of the formula under inspection can thus be calculated in a limited number of steps, provided the model is finite. The method is therefore tractable and the whole semantic approach seems promising.

[22]Cf. the example of the weather station from the general introduction.

The problem here resides in finding a characteristic model. Notice that, of course, there is no problem finding a model verifying the information, if consistent: this follows from the completeness theorems proved earlier. Apart from this relatively easy satisfaction problem there is the more severe problem that such models usually verify simply too much. So we are looking for what might be called the weakest verifying model, but then again, these models soon become quite large, sometimes even infinite. In other words, we are looking for the smallest yet weakest verifying model. Since 'weakest' implies that the model is essentially maximal and 'smallest' that it is minimal, finding such a model is a highly non-trivial task.

In the next chapters we will therefore restrict ourselves to this modeltheoretic method and choose a strong background logic, usually the modal logic **S5** or a (partial) variant of this, and only shortly discuss alternative systems, such as the multi-modal counterpart of **S5** and relaxations such as **S4**.[23] Now, at least for computer systems, a strong epistemic logic such as **S5** seems quite appropriate. Yet for the simple case of one single agent (the computer) and a strong modal logic, there are limitations to what can be modelled in this way. In particular, *incomplete* information should be of a form that allows introspection. But even when such a finite characterizing model (somewhat wishfully called a 'miniature') exists, there may be practical problems related to its size. It is argued that the latter problem can be solved by introducing partiality into the possible world models. The use of partial models for the semantic approach to knowledge-based systems will be taken up in the last chapter, chapter 9. Before doing so, we will study classical possible world models in the next chapter.

Although the application of epistemic logic to future information systems is surely promising and motivates research in this direction, we would like to emphasize that we are still far from fully realizing the program. So, what will be achieved in the final chapters may be viewed as a first modest step towards using models as knowledge-bases.

[23]Cf. chapter 4 for these modal systems.

Chapter 8

Classical models for knowledge-based systems

The leading question in this chapter[1] is: Is it possible to store knowledge in a finite possible world model and use the model as a knowledge base? We show that for somewhat idealized knowledge (captured by the **S5** logic) we can effectively find such a model, given some reasonable background assumptions concerning the sort of data involved, such as (a kind of) *introspection*. In fact several possible notions of model representation are studied and related to corresponding types of inference. We conclude with some results about extensions and relaxations of the **S5** logic, which systems are meant to capture, respectively, the knowledge of distributive systems and the knowledge of human agents.

8.1 Introduction

This chapter and the next one are devoted to the question whether we can store information in a logical model which is meant to be used as a knowledge base, so that the data are not merely encoded and inspected upon querying, but may be combined in order to derive new conclusions and thus provide a potential infinity of answers. As the class of models for a given set of data can be very large (in general even infinite), one cannot check whether something follows from the available information by inspecting virtually every model. Moreover, we would have to inspect this class of models for each query over and over again.

One prominent way to deal with this multitude is to enclose all information in one single finite model, if possible. So the central problem is:

> Given a set of data D, is there a characterizing *finite* model for D?

[1]The present chapter is an edited version of [Th92a], ©Elsevier Science Publishers 1992. An early predecessor of this chapter is the TILL paper no.111 in Dutch ('Kripke modellen voor kennisbanken I'), May 1987. The impetus to the present approach is the local interest in somewhat related work on linked data bases, see [Bu90]. The formal program executed here is sketched independently in [HV91].

As far as information is concerned which consists solely of primitive facts – logically speaking, of (atomic) propositions – and the inferences are also factual (i.e. purely propositional), there is hardly a theoretical problem, although practice can be quite difficult. On the theoretical level at which this investigation takes place, the situation gets complicated when the system not only reasons about facts, but also about itself. As we have seen in the introduction to part III, this self-reflection is not an intellectual curiosity, but an essential prerequisite to the success of building, for example, user-friendly information systems. So it should be possible to express epistemic self-reflection in the logic. The most obvious way is to use the language of epistemic logic. In this part of the book the epistemic operator K ought to be read as: 'the system knows that'. When the system has to reason not only about itself, but also about the knowledge of the user, more epistemic operators will be necessary (for instance K_u for 'user knows that' versus K_s for 'system knows that'). In this chapter we will confine ourselves mostly to the case of just one 'agent', the system itself. Apart from this we will constrain the logical language in other respects in order to avoid, for instance, problems of quantification. These restrictions allow the real problem to come out more clearly. In this way the representation language has actually become one of propositional modal logic.

The feasibility of finite model representation is connected to the strength of the chosen logical engine. Following, among others [Hi62] and [Le78], we will take knowledge to be true, and require *positive introspection* (i.e. knowing, of what one knows, that one knows it) and, unlike Hintikka and Lenzen, *negative introspection* (i.e. knowing, of what one doesn't know, that one doesn't know it) of the computer system.[2] In all, the logic will initially be the modal system **S5**. Later on, we will investigate relaxations of **S5**, such as **S4** (i.e. without negative introspection). Since our ultimate purpose is to build a many-agents system, we will make some tentative remarks on the *multi-modal* counterpart of **S5**, $\mathbf{S5_{(m)}}$.

Since we are adopting a non-syntactic approach, all these axioms and deduction rules will not be explicitly but implicitly present in the models. Model representation may even be said to avoid some problems of syntactic approaches, such as the choice of the specific logical language, the actual representation of knowledge and the way to perform deduction (cf. [Ro85]). An extreme option here is to skip the intermediate logical language entirely, and directly interpret (a subset of) natural language through the models.

8.2 Local models

The initial idea of knowledge representation by Kripke models goes as follows. A finite amount of (possibly incomplete) information is represented in a finite model ('miniature'). This model is stored in the computer system and is consulted when a question is asked which is relevant to the domain of application. A model represents

[2]Note, however, that Hintikka and Lenzen are dealing with *human* agents only, so there there is no disagreement here.

the information D, we suppose, when it verifies D, but doesn't verify anything that doesn't follow from D, in other words, when all and only consequences of D are verified. Here D is obviously a finite set of well-formed formulas of the appropriate epistemic language $\mathcal{L}_K(Prop)$, where *Prop* is also finite.[3] Usually the evaluation is considered to start from a fixed world w. We will then speak of a *local* model $\langle M, w\rangle$. M is a standard Kripke model $\langle W, R, V\rangle$, as described in sections 4.6.1.

Definition 8.1 *$\langle M, w\rangle$ is a* **local miniature** *for D iff M is finite and for every φ: $M, w \models \varphi \Leftrightarrow D \models \varphi$.*

In checking whether some structure is a (local) miniature, it turns out to be convenient to break down the condition of exactly verifying all D-consequences into several parts (equivalent to the above definition):

Proposition 8.1 *$\langle M, w\rangle$ is a local miniature for D iff*

- *M is finite,*
- $M, w \models D$,
- $M, w \models \varphi \Rightarrow D \models \varphi$.

Proof: directly from definition 8.1 and the definition of logical consequence. ■

On which (syntactic) conditions does a local miniature exist? *Consistency* of D is not sufficient: for example, $K(p \vee q)$ is consistent, but does not have a local miniature: if $M, w \models K(p \vee q)$ then $\langle M, w\rangle$ verifies p and q (or p and $\neg q$, or $\neg p$ and q), but neither p nor q (nor $\neg p$ nor $\neg q$) is a consequence of $K(p \vee q)$. In fact every incomplete set of data will be problematic for virtually the same reason.

Theorem 8.1 *D has a local miniature iff D is both complete and consistent.*[4]

Proof:
($\Rightarrow$) Because of the completeness theorem, the existence of the miniature $\langle M, w\rangle$ implies consistency and completeness of D: $D \nvdash \varphi \Leftrightarrow D \not\models \varphi \Leftrightarrow M, w \not\models \varphi \Leftrightarrow M, w \models \neg\varphi \Leftrightarrow D \models \neg\varphi \Leftrightarrow D \vdash \neg\varphi$.
($\Leftarrow$) If D is consistent, it has a local model that can be converted into a finite model $\langle M, w\rangle$ for D by a suitable filtration over the elements of D and their subformulae.[5] Suppose $D \not\models \varphi$, then, since D is complete, $D \models \neg\varphi$, so $M, w \models \neg\varphi$ and consequently $M, w \not\models \varphi$. From proposition 8.1 it follows that $\langle M, w\rangle$ is a miniature. ■

Notice that the nature of the underlying logic is hardly of importance to the proof.

---

[3] See e.g. section 5.3

[4] Recall that D is complete (in this sense) if $D \vdash \varphi$ or $D \vdash \neg\varphi$. D is complete and consistent is equivalent to: $\overline{D}$ is maximally consistent (where $\overline{D} = \{\varphi \mid D \vdash \varphi\}$).

[5] See [Ch80] or [HC84].

The theorem holds both for **S5** and for richer modal systems such as $\mathbf{S5_{(m)}}$, but also for weaker systems, such as **S4**, just as long as the logic has the finite model property (FMP).[6] The finiteness of D is also of vital importance to the proof.

As incomplete information implies logical incompleteness, we may conclude at this point that local miniatures are unfit for our purpose. In practice the system will, of course, seldom have complete information at its disposal. Therefore we will continue our quest for an adequate notion of characterizing model.

8.3 Global models

A local miniature enforces complete knowledge, essentially because each formula is either true or false in the designated world w. In order to avoid the problem, formulas have to be evaluated in other worlds than w too. Let φ be (globally) true on a model M iff for all worlds w in M: $M, w \models \varphi$ (notation: $M \models \varphi$). This leads to a different notion of miniature:

Definition 8.2 *M is a* **global miniature** *for D iff M is finite and $M \models \varphi \Leftrightarrow D \models \varphi$ for every φ.*[7]

Here too, an equivalent but possibly more intuitive form exists:

Proposition 8.2 *M is a global miniature for D iff*

- *M is finite,*
- $M \models D$,
- $M \models \varphi \Rightarrow D \models \varphi$.

Observe that the condition of finiteness is essential, both for practical purposes and for making the search for global miniatures an interesting puzzle. Without finiteness the following method leads to a (possibly infinite) global 'miniature' for information of essentially the form $K\alpha$, where α is an arbitrary formula. Assume the modal logic to be (a normal extension of) **S4**, i.e. we may suppose the accessibility relation to be reflexive and transitive.

Consider the set Γ of all *non-consequences* of $K\alpha$. For each $\gamma \in \Gamma$, let M_γ be a (smallest) counterexample to $K\alpha \models \gamma$ (for logics with the FMP such a counter-model will be finite). Notice that by transitivity and the generation lemma, the counter-model may be chosen to be generated from a single world in one step.[8] Now take the disjoint union[9] of all such M_γ. Then, apart from finiteness, the union model M meets the other

[6] A system has the FMP if it is characterized by a class of *finite* models. That $\mathbf{S5_{(m)}}$ has the FMP can be shown by simply multiplying the filtration for **S5**. Cf. [Ch80], [HC84] and section 4.6.2 here.

[7] Or, $\text{TH}(M) = \overline{D}$, with $\text{TH}(M) = \{\varphi \mid M \models \varphi\}$.

[8] See section 2.3.

[9] Informally speaking, the operation of disjoint union amounts to putting different models into one single model, *without* interconnecting them. Cf. definition 8.3 for a formal statement.

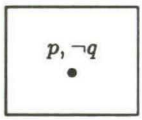
conditions of proposition 8.2: since every generated submodel locally verifies $K\alpha$, it verifies α in all its generated submodels ('components'), and so it globally verifies $K\alpha$. The third requirement is met since, by contraposition: if φ is not a consequence of $K\alpha$, then φ is locally falsified in M_φ, and so, again by the generation lemma, locally falsified in M.

Notice this argument can be simplified and, at he same time, generalized to arbitrary normal systems, once we move to a relation of *global consequence*.[10] Let $D \mathrel{|\!\approx} \varphi$ denote global consequence, i.e. every global model of D also globally verifies φ. Now let M be the disjoint union of counterexamples to $D \mathrel{|\!\approx} \gamma$ for all γ. Then obviously $M \models D$ and $M \models \varphi \Rightarrow D \mathrel{|\!\approx} \varphi$. So, we can always obtain a characteristic global model with respect to global consequences. Caveat: this involves a different notion of 'miniature'. And, once again, finiteness is not guaranteed.

Returning to the original notion of global miniature, we would like to find a condition that determines when a given set of data is characterized by a miniature. To get an idea of the problems at stake, we first give a number of concrete examples. Until further notice the modal logic will be **S5**, so accessibility in the models is an equivalence relation. [11]

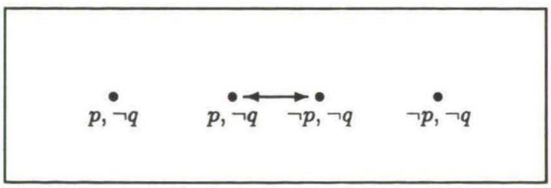
First we shall give an example of a miniature for *complete* information, to contrast with later examples of incomplete information.

Example 8.1 *Take Prop* $= \{p, q\}$ *and* $D = \{Kp, K\neg q\}$. *The minimal global (and local) miniature* M *for* D *is:*



The next case concerns a simple form of incomplete information.

Example 8.2 *Take Prop* $= \{p, q\}$ *and* $D = \{K\neg q\}$. *A global miniature* M *for* D *will typically be of the following form:*



This miniature is inspired by the equivalence

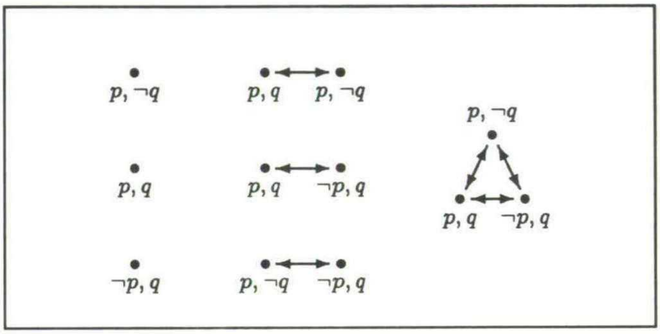
$$K\neg q \leftrightarrow (Kp \wedge K\neg q) \vee (\neg Kp \wedge \neg K\neg p \wedge K\neg q) \vee (K\neg p \wedge K\neg q)$$

[10]The heuristics of this method was suggested by Johan van Benthem.

[11]For different w and v, if both wRv and vRw, then there is a double arrow between w and v in the diagrams; if wRv but not vRw (e.g. for non-symmetric **S4** models), there is a single arrow from w to v. Looping arrows which are redundant by virtue of reflexivity (as for **S4** and **S5** models), have been left out.

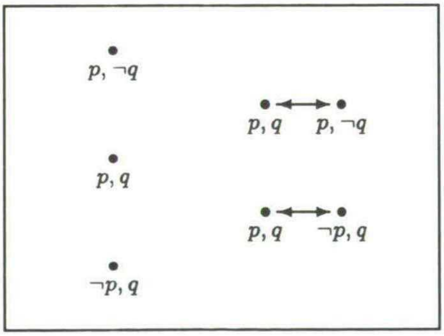
in which case $\vee$ amounts to an exclusive disjunction. Here, each of the disjuncts has a characterizing model which corresponds to a component of M. Moreover, the division into several components is inevitable when not all propositional variables occur in D. As a matter of fact such a scattered distribution also can be found in cases where the information is incomplete in another sense:

Example 8.3 *Take* $D = \{K(p \vee q)\}$ *and* $Prop = \{p, q\}$. *The global* **S5** *miniature for this information is*



The incomplete knowledge occurring in example 8.3 is of an entirely different nature than the one in the next interesting case:

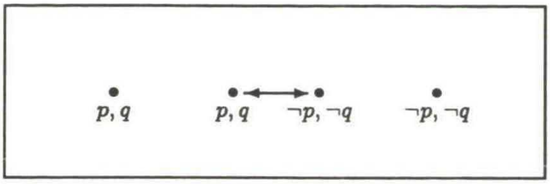
Example 8.4 *Again take* $Prop = \{p, q\}$ *but now* $D = \{Kp \vee Kq\}$. *D has the following* **S5** *miniature:*



This result seems diametrically opposed to the suggestion which looms up from [HM85] that $Kp \vee Kq$ has no characterizing model (but see sections 8.5 and 8.6 for further analysis).

A final example of a global miniature for a stronger kind of incomplete information is:

Example 8.5 *Let* $Prop = \{p, q\}$ *and* $D = \{K(p \leftrightarrow q)\}$. *A global miniature for D is:*



What is the syntactic criterion for the existence of a global miniature? Apart from mere consistency the alternative of *global consistency*, i.e. existence of a non-empty verifying model, comes to mind. Contrary to the local case we note that here global consistency is neither necessary nor sufficient: the formula $(p \wedge \neg Kp) \vee Kq$ is globally consistent (i.e. it has a non-empty global model), but it does not have a global miniature. For suppose that N is a global miniature for $D = \{(p \wedge \neg Kp) \vee Kq\}$. Then (i) $N \models (p \wedge \neg Kp) \vee Kq$, and so $N \models Kq$. For suppose not, i.e. for some w (ii) $N, w \not\models Kq$, then, by (i), $N, w \models p \wedge \neg Kp$, and thus for some v such that wRv: $N, v \not\models p$, and again by (i), $N, v \models Kq$, hence by euclidicity $N, w \models Kq$, contradicting (ii). Therefore $N \models Kq$, yet $D \not\models Kq$ (the left disjunct can be verified locally without q being true). Note that although global consistency is of a semantic nature, it can easily be transposed into a purely syntactic formulation (in the sequel $K[D] = \{K\varphi \mid \varphi \in D\}$):

Proposition 8.3 *D is globally consistent iff $K[D]$ is consistent.*

Proof: The right-hand condition is clearly necessary, but also sufficient, for suppose $K[D]$ is consistent, then there is an **S5** model M, such that for any $\varphi \in D$: $M, w \models K\varphi$, thus for the generated submodel[12] $M_w \models \varphi$, since it can be shown by induction that for every ψ and u such that $wRu : M, u \models \psi \Leftrightarrow M_w, u \models \psi$ (by the generation lemma, cf. section 2.3). ■

So consistency of $K[D]$ does not suffice, in fact it is not even a necessary condition for the existence of a miniature! The point is that the definition of *global miniature* allows of *empty* models. The empty model verifies every formula vacuously, and so may be used to model inconsistent information. But proposition 8.3 may be strengthened to the related condition that everything in $K[D]$ follows from D, or equivalently $D \models \varphi \Rightarrow D \models K\varphi$, which amounts to the plausible requirement that the system knows the data, i.e. a kind of introspection. We arrive at the main result of this chapter:

[12]Recall that M_w is M restricted to the worlds accessible from w.

Theorem 8.2 *D has a global miniature iff* $D \vdash \bigwedge\!\!\!\bigwedge K[D]$. [13]

Proof: this is essentially the heuristic argument given in the beginning of this section, now using logical finiteness of **S5**.[14]

($\Rightarrow$) The condition is clearly necessary, since if D has a global miniature M, then for every $\varphi \in D$: $D \models \varphi \Rightarrow M \models \varphi \Rightarrow M \models K\varphi \Rightarrow D \models K\varphi$. Therefore $D \models \bigwedge\!\!\!\bigwedge K[D]$, thus by the completeness theorem $D \vdash \bigwedge\!\!\!\bigwedge K[D]$.

($\Leftarrow$) Let Φ be a set of representatives of equivalence classes of formulas induced by the relation of logical equivalence. Because of the finiteness of *Prop* and the logical finiteness of **S5**, Φ is also finite. Now let M be the disjoint union of the smallest counterexamples to non-consequences of D: (*vide* definition 8.3 below for details)

$$M = \bigoplus\{N \mid N \models D, N \text{ is tight and reduced, and } N \not\models \varphi \text{ for some } \varphi \in \Phi\}$$

Then M is a global miniature according to proposition 8.2:

- M is finite because Φ is finite and each N is too (because it contains no more than 2^n worlds when $|Prop| = n$);
- $M \models D$;
- if $D \not\models \varphi$ then there exists a $\psi \in \Phi$ such that $\models \varphi \leftrightarrow \psi$, ergo $D \not\models \psi$. So there is a model $\langle L, w\rangle$ such that $L, w \models D$ and $L, w \not\models \psi$. Because $D \models K[D]$, $L, w \models K[D]$ holds too, and so $L_w \models D$ and $L_w \not\models \psi$. L_w is tight, and reduction supplies the N for which $N \models D$ and $N \not\models \varphi$, and this is a submodel of M, so also $M \not\models \varphi$. ∎

We used the following notions here:

Definition 8.3 *(with* $M = \langle W, R, V\rangle$ *and* $N_i = \langle W_i, R_i, V_i\rangle$*):*

1. *M is* **tight** *(and R is* **universal***) iff* wRu *for every* $w, u \in W$*;*
2. *M is* **reduced** *iff* $V(p, w) = V(p, v)$ *for each* $p \in Prop \Rightarrow w = v$*;*
3. $M = \bigoplus\{N_i \mid i \in I\}$ *(the* **disjoint union** *of the* N_i*'s) where:*
   - $W = \{\langle w, i\rangle \mid w \in W_i\}$,
   - $\langle w, i\rangle R\langle u, j\rangle \Leftrightarrow wR_iu \wedge i = j$, *and finally*
   - $V(p, \langle w, i\rangle) = V_i(p, w)$.

Explanatory note: 'tightness' guarantees the cohesion of a model structure; the definition given here is restricted to the (simple) **S5** case.[15] 'Reduction' is a purely technical operation which removes superfluous worlds. Disjoint union[16] combines a number of models as separate components into a new model; a component of a model is a tight submodel.

Given the introspection property for D, the proof of theorem 8.2 licenses the following procedure for generating miniatures:

[13]Recall that $\bigwedge\!\!\!\bigwedge$ indicates the finite conjunction over the elements of $K[D]$: if $D = \{\delta_1, \dots, \delta_N\}$ then $\bigwedge\!\!\!\bigwedge K[D] = K\delta_1 \wedge \cdots \wedge K\delta_N$.

[14]Logical finiteness means that, given a finite *Prop*, there are only finitely many different formulas, up to equivalence. **S5** is logically finite, but **S4** and $\mathbf{S5}_{(\mathbf{m})}$ are not. Cf. section 4.6.3

[15]cf. [HC84].

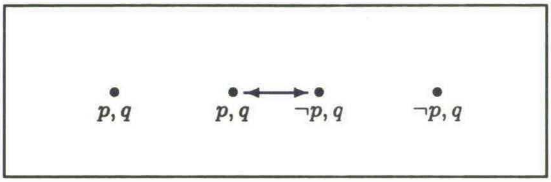
[16]See also corollary 2.3.

(DU1) create the set of all tight reduced models (there are a finite number of them, if $|Prop| = n$ at most 2^{2^n});

(DU2) take the *disjoint* union of the D models out of this set.

With this procedure examples 8.2-8.4 are easily constructed. Moreover, theorem 8.2 shows that the procedure does in fact produce a model which is not a miniature when D is not introspective:

Example 8.6 *Let $Prop = \{p, q\}$ and $D = \{(p \wedge \neg Kp) \vee Kq\}$. The construction* DU *results in a model that is not a global miniature for D:*



If one also wishes the procedure DU to succeed when the system is not introspective with respect to D, one has to switch to a *global consequence* relation.

8.4 Global consequence and coarse miniatures

From the initial considerations in the previous section the suggestion looms that perhaps one can avoid the condition of introspection on D by using a global consequence relation, whilst the notion of 'miniature' is accordingly adjusted. We will now elaborate on this point.

Definition 8.4 (global consequence)
$D \approx\!\!\!\!\models \varphi$ iff $M \models D \Rightarrow M \models \varphi$ for every model M.[17]

In many cases it is possible to obtain the just defined notion from the usual relation of logical consequence by sufficiently augmenting D. In example 8.6, Kq was not a standard consequence of $(p \wedge \neg Kp) \vee Kq$, but it is a global consequence thereof, and, perhaps surprisingly, an ordinary consequence of $K((p \wedge \neg Kp) \vee Kq)$, because in **S5** this formula is equivalent to $K(p \wedge \neg Kp) \vee Kq$.[18] The first disjunct of the last formula contradicts the **S5**-axioms, so Kq indeed follows now. This suggests a more general connection:[19]

[17] Cf. [vB85], pp. 37/38.

[18] *Vide* [HC68], p.51, theorem 29.

[19] In fact, the proposition holds not only for **S5**, but for every extension of **S4**. On its turn, this relation is generalized in [FHV90] to arbitrary normal systems, provided we introduce the modality K^+ ('common' knowledge for one agent!), which is interpreted by means of the transitive closure R^+ of R. Then we obtain $D \approx\!\!\!\!\models \varphi \Leftrightarrow K^+[D] \models K^+\varphi$, in accordance with an observation in [GP89].

Proposition 8.4 $D \approx\!\!\!\!\models \varphi \Leftrightarrow K[D] \models \varphi$.

Proof: in analogy with proposition 8.3 ■

The notion of global consequence induces a new type of miniature:

Definition 8.5 *M is a* **coarse miniature** *for D if M is finite and* $M \models \varphi \Leftrightarrow D \approx\!\!\!\!\models \varphi$ *for each formula* φ.

The notion of *course miniature* clearly extends that of *global miniature*, for any global miniature is also a coarse miniature. The reverse only holds if the information is introspective. Yet another way to relate the two notions is:

Proposition 8.5
M is a coarse miniature for D iff M is a global miniature for $K[D]$.

Proof: follows directly from proposition 8.4 and definitions 8.2 and 8.5. ■

Theorem 8.3 *Every D has a coarse miniature.*

Proof: notice that $K[D]$ is introspective and apply theorem 8.2 and proposition 8.5. ■

Because the notion 'global consequence' as well as 'coarse miniature' can be reduced to *standard consequence* and *global miniature* according to propositions 8.4 and 8.5, these notions are not of importance for **S4** and its extensions, but could, in principle, lead to different results for weaker modal systems.

8.5 Tight miniatures

With example 8.4 we noted that Halpern&Moses had come up with a seemingly opposite result for $Kp \vee Kq$ in [HM85]. One of the causes of this difference is that they only consider *tight* models. With this restriction the following procedure (TU) leads to the greatest reduced tight model, if it exists:

(TU1) take the union of the tight models for D and choose the accessibility relation to be universal (the 'total union');[20]

(TU2) reduce the united structure, i.e. identify worlds in the total union model that assign the same truth-value to each propositional variable;

(TU3) check whether D is still true in this tight structure.

[20] This amounts to dropping the accessibility relation from both the models and the truth conditions.

(The total union of the models M and N will be denoted by $M \uplus N$, the disjoint union by $M \oplus N$).

Step TU3 fails with example 8.4: $Kp \vee Kq$ has no greatest reduced tight model. The union model verifies $\neg Kp \wedge \neg Kq$, so cannot be a model for D.

Although the notion of a greatest reduced tight model is quite different from that of a tight miniature, the latter is worth investigating. We restrict the notion of 'tight miniature' to 'tight global miniature', as (again by the generation lemma) a tight local miniature exists precisely when there is a local miniature.

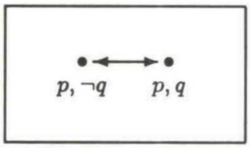
Because the notion of 'tight miniature' is stronger than that of 'global miniature', and resembles 'local miniature' in other ways too, we may wonder whether conformity goes even further. This is not the case: it is particularly not so that a tight miniature characterizes a maximally consistent set, or *vice versa*; in fact the equivalence (cf. theorem 8.1) does not hold in either direction: (D is finite, as always)

D has a tight miniature $\not\Leftrightarrow$ D is complete and consistent.

Proof:

($\not\Leftarrow$) if *Prop* $= \{p\}$ then (by the Lindenbaum lemma and logical finiteness) there is a finite D such that $\{p, \neg Kp\} \subseteq D$, where $\overline{D}$ is maximally consistent — for $\{p, \neg Kp\}$ is (locally!) consistent. But D surely has no global model (and so: no tight one either).

($\not\Rightarrow$) for the invalidity of the converse, consider the tight model M:



By logical finiteness $\mathrm{TH}(M)$ is equivalent to some finite D, e.g. $\{Kp, \neg Kq, \neg K\neg q\}$. It is easy to check that M is a global miniature for D. Although D is consistent, it is not theoretically complete: neither $M \models q$ nor $M \models \neg q$, so, because M is a miniature for D: $D \not\vdash q$ and $D \not\vdash \neg q$. ∎

A syntactic characterization of this new notion after the style of theorem 8.2 is:

Theorem 8.4 *D has a tight miniature iff D is introspective ($D \vdash \bigwedge\!\!\!\!\bigwedge K[D]$) and K-complete (for all φ : $D \vdash K\varphi$ or $D \vdash \neg K\varphi$).*

Instead of proving theorem 8.4 directly, we will first show a connection between the modal conditions mentioned in the theorem and a notion introduced by Stalnaker.

Definition 8.6 *A set T of formulas is* **(S5–)stable** *iff*

1. *T contains the pL axioms;*

2. *T is deductively closed: if $\varphi \in T$ and $\varphi \to \psi \in T$ then $\psi \in T$;*

3. *if $\varphi \in T$ then $K\varphi \in T$;*

4. *if* $\varphi \notin T$ *then* $\neg K\varphi \in T$;

5. T *is consistent with respect to* pL.[21]

In 3 and 4 we can replace the conclusion by an equivalence, because of 5. To prove theorem 8.4 it suffices to remark that the conditions for tightness of the miniature can be reformulated in terms of stability:

Lemma 8.1 *D is consistent, K-complete and introspective iff $\overline{D}$ is stable.*

Proof: straightforward ■

Now tight models and stable sets are intimately related:

Theorem 8.5 *T is stable $\Leftrightarrow$ for some tight M:* $\text{TH}(M) = T$.

Proof: [HM85, lemma 2 and proposition 3] ■

Proof of theorem 8.4: as before we may suppose D to be consistent. If D is introspective and K-complete, then (by lemma 8.1 here) $\overline{D}$ is stable, and so, by theorem 8.5, there is an M such that $\text{TH}(M) = \overline{D}$. By [HM85, proposition 3] M is reduced and by the logical finiteness of **S5**, M must be finite. Consequently M is a tight global miniature for D. The proof of the converse is quite simple: let M be a tight global miniature for D, then (theorem 8.2) D is introspective. To prove K-completeness, assume that $D \nvdash K\varphi$, then *not* $M \models K\varphi$, i.e. for some v and u such that vRu: $M, u \not\models \varphi$. But since M is tight, u is accessible from every world in M, thus $M \models \neg K\varphi$, and so $D \vdash \neg K\varphi$. ■

Although theorem 8.4 shows that the conditions of existence of a tight miniature hardly differ in strength from those of a local miniature, the former can be fulfilled by invoking a modal version of the 'closed world assumption' (CWA). D is augmented, resulting in D', by the clauses:

- $\varphi \in D \Rightarrow K\varphi \in D'$;[22]
- $D \nvdash \varphi \Rightarrow \neg K\varphi \in D'$.

(this is again effectively possible by the logical finiteness of **S5**), after which procedure TU produces the tight miniature, provided D' is still consistent. In doing so we have implicitly met the criterion of 'honesty' of the system's knowledge, which is dealt with in the next section.

[21] The use of the letter 'T' stems from the fact that 1 and 2 imply that a stable set is a *theory*: i.e. if $T \vdash \varphi$ then $\varphi \in T$. Moreover, $\vdash$ does not have to be restricted to propositional inference, since we can show that the axioms of **S5** are contained in a stable T, which, by 3, is closed under **N**. Here is a sample derivation of the fact that $\mathrm{T} \in T$: $\varphi \in T$ or $\varphi \notin T$, so $\varphi \in T$ or $\neg K\varphi \in T$, then since $\varphi \rightarrow (K\varphi \rightarrow \varphi)$ and $\neg K\varphi \rightarrow (K\varphi \rightarrow \varphi)$ are propositionally valid, thus (by 1 and 2) contained in T, we obtain, again by 2, that $K\varphi \rightarrow \varphi \in T$.

[22] If D is introspective we may omit K from this clause; in that case D' amounts to an extension of D.

8.6 Honesty

Which D are permitted, if the system *only* knows D? For many D such *circumscription* is not possible without stumbling into inconsistencies. Example 8.4 produces a paradigm: if the system *only* knows $Kp \vee Kq$, then $\neg Kp$ and $\neg Kq$, and consequently $\neg Kp \wedge \neg Kq$, which contradicts $Kp \vee Kq$! This is why such a D is called *dishonest*: one cannot sincerely claim to know p or to know q without actually knowing at least one of the two. Or, closer to natural language, it seems self-contradictory to say you *only know whether p*. For proper understanding, a couple of remarks must be made here.

In the first place we have to distinguish such dishonest knowledge from the case in which the system only knows that p or q, in other words, when D only contains the information $K(p \vee q)$ (cf. examples 8.3 and 8.4). The latter case clearly does not involve dishonest knowledge: $K(p \vee q)$ is consistent, even when combined with $\neg Kp$, $\neg Kq$, $\neg K\neg p$ and $\neg K\neg q$.

In the second place, although $Kp \vee Kq$ seems unfit to be plain information, it is proper as the representation of a question (query) or answer. This is reflected in the fact that 'dishonest' knowledge may become proper when embedded: an agent (human being or computer system) can only know of another agent that it knows p or that it knows q. So, although $K(Kp \vee Kq)$ is not honest, $K_a(K_b p \vee K_b q)$ is.

This is one of the reasons for favouring our more general approach of miniatures: in a multi-modal setting, dishonest formulas can become honest when embedded, so the 'dishonest' miniature may occur as a substructure of the whole 'honest' miniature.

Several formal criteria for honesty are proposed in [HM85], which are then shown to be equivalent. We will present two of the five criteria here.

Definition 8.7 *D is* **stable-honest** *if there exists a stable T which contains D, and $T \cap pL$ is a unique minimal set (for D).*

This definition is justified by the observation of [Mo85] that every stable T is completely determined by $T \cap pL$, the set of K-free formulas in T. Minimality is prompted by 'not knowing more than D'. It is shown that a stable set cannot be strictly contained in another one, therefore we cannot demand minimality of T itself. That unicity is required may be seen by reconsidering $Kp \vee Kq$; notice that either p or q is in its stable extension (otherwise we obtain $\neg Kp \wedge \neg Kq$) — so there is no unique minimum. [23]

Definition 8.8 *D is* **model-honest** *if the total union of the D models is a model for D.*

Fortunately these two seemingly different notions of 'honesty' coincide.

Theorem 8.6 (Halpern&Moses) *D is stable-honest iff D is model-honest.*

[23] The terminology used here is slightly different from that in [HM85], where unicity is probably implied by minimality.

Now what is the connection between their approach and ours?

In the first place: their **S5** models are much simpler, so why "do it the hard way" with structured models? The reason for this step lies in the possibility of *recursive extension*. Although we have not investigated these cases in full detail yet, it is *a priori* clear that a multi-modal logic (with several agents), and also the unimodal (one agent) **S4** logic cannot do without accessibility relations in their models. So it seems advantageous to treat the simple **S5** case on a par.

In the second place we note that, quite different from what one might expect, the union model is in general *not* a tight miniature, but will turn up as the largest component of the global miniature, in the sense that all components will be submodels of this largest component.

Theorem 8.7 *D is honest $\Leftrightarrow$ D has a coarse miniature that is closed under total union of the submodels.*

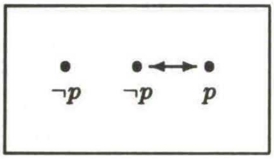
Proof:
($\Rightarrow$) By proposition 8.5, D always has a coarse miniature which coincides with the global miniature M for $K[D]$. Let $M = \bigoplus\{N_i : N_i \models K[D]\}$ (cf. the proof of theorem 8.2). Now if D is model-honest, M will be closed under total union of components, for the N_i are both global models for $K[D]$ and for D.
($\Leftarrow$) Suppose M is a global miniature for $K[D]$ that is closed under $\uplus$ of components. Assume moreover that D is globally consistent (otherwise it is trivially model-honest). Then there must be a greatest component N. We now prove that N is the greatest D model. Notice that, of course, $N \models D$. It remains to be proven that each (tight, reduced) D model is contained in N. Choose a tight and reduced model L such that $L \models D$. Again we may suppose L to be non-empty. L will be finite, so $W_L = \{w_1, \ldots, w_k\}$. Let *Prop* be $\{p_1, \ldots, p_n\}$. Now define, for each world $w_i \in W_L$, a formula σ_i that characterizes the state of affairs in w_i. First we introduce the auxiliary $\sigma_{i,j}$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n$.

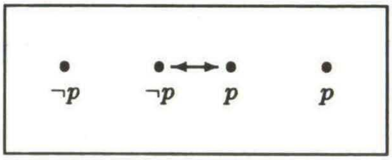
$$\sigma_{i,j} = \begin{cases} p_j & \text{if } V_L(p_j, w_i) = 1 \\ \neg p_j & \text{if } V_L(p_j, w_i) = 0 \end{cases}$$
$$\sigma_i = \sigma_{i,1} \wedge \ldots \wedge \sigma_{i,n}$$
$$\sigma = \neg K \neg \sigma_1 \wedge \ldots \wedge \neg K \neg \sigma_k$$

Then for every i: $L, w_i \models \sigma_i$ and so $L \models \sigma$. Now suppose $L \not\subseteq N$ then for each component C of M: $L \not\subseteq C \Rightarrow$ for each such C there is an $i \in \{1, \ldots, k\}$ such that for all worlds w in C there is a $p \in$ *Prop* for which $V_L(p, w_i) \neq V(p, w)$. So for any such C and w: $C, w \not\models \sigma_i$ holds, therefore $C \models K\neg\sigma_i$, and so $C \models \neg\sigma$. Thus $M \models \neg\sigma$, hence $D \models \neg\sigma$. This, however, contradicts $L \models D \cup \{\sigma\}$. ■

So, another way to express the same result would be that D is honest iff $K[D]$ has a global miniature with a maximal component. Notice that the second part of the proof of theorem 8.7 demonstrates another important fact about miniatures, viz. their uniqueness. Theorem 8.7 also shows that the notion of 'honesty', too, can be expressed in our framework. In other words, our theory will call the same sets of data 'honest' as [HM85]. It is a totally different matter whether honest information in fact fits the intuition of determining a knowledge state.

A borderline case with respect to honesty is the information $D = \{\neg Kp\}$. We could surmise that this example is like the standard counterexample $Kp \vee Kq$, because it contains a comparable kind of indeterminacy. This connection seems to be supported by the equivalence $\neg Kp \leftrightarrow K\neg p \vee K(\neg Kp \wedge \neg K\neg p)$. But even so, $\neg Kp$ differs from $Kp \vee Kq$ in that it *is* in fact honest: the stable extension of D with a uniquely smallest subset of pL formulas is the stable set which only contains purely propositional formulas that are tautologies. The greatest tight model of D is then also the greatest possible tight model for this *Prop*. Notice however that by means of our miniatures we can discriminate various types of very weak knowledge. For example, the above $D = \{\neg Kp\}$ and total lack of information ($D = \emptyset$) have different miniatures, even when $Prop = \{p\}$:

D: 

$\emptyset$: 

The left-hand model characterizes the information of "not knowing p", whereas the 'no information model' on the right-hand side characterizes full absence of information about p. This distinction is useful, above all, from a dynamic point of view: the set $\emptyset$ characterizes complete lack of information *at a certain point of time*, and can be consistently extended by the information Kp at some later moment. For D this is not possible without contradiction. One can imagine this type of information to occur in practice in, for example, an automatized weather station: if a measuring-device is out of order or does not provide a decisive answer, the central data processor has to register this as not knowing the dimension in question (for instance temperature); compare the opposite situation where the data from this measuring-device have not yet arrived at the central computer.

8.7 Dimensions for extension and variation

Various dimensions have been mentioned above:

1. The richness of the representation language
2. The logical power of the modal system
3. The kind of semantic structures.

We can vary one or more of these parameters and study their representational possibilities. We shall discuss the different dimensions below.

8.7.1 Richness of the logical language

The syntax can be extended in various ways, for example:

- with predicates, variables and quantifiers (quantified modal logic);
- by introducing more agents (multi-modal logic).

[Le84b] deals with quantification, even though, in a strict sense, it does not deal with *knowledge* but with uncertain information, i.e. *belief*. The truth axiom **T** is then replaced by the consistency axiom **D**: $\neg(B\varphi \wedge B\neg\varphi)$.

Introduction of multiple agents is immanent to studies of distributed systems, where a number of interconnected processors exchange information. For this purpose a homogeneous logic in which the processors use the same modal system is adequate. As noticed in [HM85] even $\mathbf{S5_{(m)}}$, multiple **S5**, poses many problems. Some of these are surely solvable (we noticed, for example, that theorem 8.1 holds for $\mathbf{S5_{(m)}}$); others are more difficult (for example defining *minimality* for $\mathbf{S5_{(m)}}$-stable sets). In fact we claim that, for the $\mathbf{S5_{(m)}}$-logic, a consistent D does not possess a global miniature, unless D is at least introspective with respect to *common knowledge* (i.e. i's knowledge of j's knowledge of ..., *ad infinitum*). Strictly speaking, this involves enriching both the logical language with an extra operator and the logical system with axioms for common knowledge, beyond $\mathbf{S5_{(m)}}$. Without common knowledge, there does not seem to be much hope for a positive result. According to the following conjecture only inconsistent data have an $\mathbf{S5_{(m)}}$ miniature, viz. the empty model.

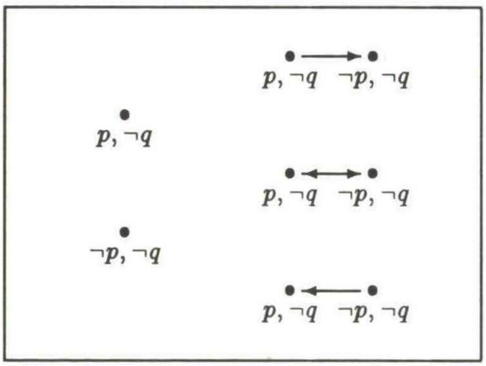
Conjecture 8.1 *For* $\mathbf{S5_{(m)}}$ *consistent D do not possess global miniatures.*

Proof: Let D be a consistent set of data. Suppose D has an $\mathbf{S5_{(m)}}$ miniature M. Either D is valid or it is not.
If D is valid, then $M \models \varphi \Leftrightarrow \models \varphi$, i.e. M verifies all and only valid formulas. This, however, is impossible: every finite model globally verifies some nonvalid formula, [FHV91, thm. 4.9].
If D is invalid, then D must be common knowledge in M. To see this let $\alpha = \bigwedge D$, $K\varphi = K_1\varphi \wedge \cdots \wedge K_m\varphi$, $K^{n+1}\varphi = KK^n\varphi$ for all n. Then by assumption $M \models \alpha$, and so by induction $M \models K^n\alpha$ for all n, and therefore $\alpha \models K^n\alpha$ for all n, where α is nonvalid. This, we claim, is impossible. For, since α is neither valid nor inconsistent, there are model N and N' such that $N, w \models \alpha$ and $N', w' \models \neg\alpha$. In fact the maximal distances $D_N(w, v)$ and $D_{N'}(w', v')$ (where the distance $D_M(x, y)$ between vertices x and y is defined as the length of the minimal path from x to y in M) for worlds v in N and v' in N' may be chosen not to exceed the modal depth d of α. Now let u be the endpoint of a longest path in N, and u' similarly in N'. Next link u and u' by $2d$ alternating edges: $uR_1u_1, u_1R_2u_2, u_2R_1u_3, \cdots, u_{2d-1}R_2u'$. Then in the resulting model L: $L, w \models \alpha \wedge \neg K^n\alpha$, where $n = D_N(w, u) + 2d + D_{N'}(w', u')$, thus contradicting common knowledge of α. ■

8.7.2 Varying the modal system

The required logical power may exceed **S5** for distributed systems, but mostly one will have to inspect *weaker* modal systems than **S5**. After all, people are not consistent in their reasoning with knowledge and not logically omniscient either. For the time being, the first modal logic to deal with will be **S4**, the proper logic for human knowledge according to [Hi62] and [Le78].

As we noticed in section 8.2, **S4** has the finite model property, and so theorem 8.1 holds for **S4** as well. So, certain types of complete knowledge can be captured by local models. Incomplete knowledge therefore requires global models. At this point we may wonder whether theorem 8.2 carries over to **S4**-knowledge: is the introspection property $D \vdash \bigwedge K[D]$ a necessary and sufficient condition for the existence of a global miniature? Inspection of the proof of theorem 8.2 however indicates that the logical finiteness of the background logic is essential. Unfortunately, **S4** is not logically finite[24], because worlds occurring in the same component which are identical as far as truth assignment to propositional variables is concerned, cannot in general be identified, since different alternatives may be accessible from them. As a consequence of this, there is no limit to the size of components in a characterizing model, so we cannot confine ourselves to finite miniatures. Although the condition of globality should suppress the complexity of the models, it appears to be difficult to supply an analogous **S4** miniature for example 8.2. For instance the asymmetric generalization of the **S5** miniature for $\{K\neg q\}$ does not suffice here:



In effect the complete lack of information about p prevents the existence of an **S4** miniature. We can prove this by using the list of formulas mentioned in note 24. First we define these formulas inductively: [25]

$$\begin{cases} \varphi_0 = \top; \\ \varphi_1 = p; \\ \varphi_{n+2} = p \wedge \neg K(p \vee K\neg\varphi_n). \end{cases}$$

For $n > 0$ neither these formulas, nor their negations follow from $D = \{K\neg q\}$. Suppose that M is the **S4** miniature of D, then, for each n, $\neg\varphi_n$ has to be refuted in a

[24] Even when $|Prop| = 1$, **S4** is not logically finite. For consider the list of formulas: p, $p \wedge \neg Kp$, $p \wedge \neg K(p \vee K\neg p)$, $p \wedge \neg K(p \vee K(\neg p \vee Kp))$,... Semantically the non-equivalence of the formulas in question becomes evident because they are discerned by a model containing chains of sufficient length.

[25] Perhaps more perspicuous equivalents are: $\varphi_{n+2} = p \wedge \hat{K}(\neg p \wedge \hat{K}\varphi_n)$, where $\hat{K} = \neg K \neg$, K's dual, or $\varphi_{n+2} = \neg(p \rightarrow K(\neg p \rightarrow K\neg\varphi_n))$.

component M_n of M. This is not entirely sufficient to prove the statement, because M_n could be the component

$$\boxed{\begin{array}{ccc}\bullet & \longleftrightarrow & \bullet \\ p, \neg q & & \neg p, \neg q\end{array}}$$

which refutes all formulas from the left world. This imperfection can be easily remedied by considering the formula $\neg\varphi_n \vee \varphi_{n+2}$ instead of $\neg\varphi_n$. A countermodel of the new formula must contain, for each n, a directed chain

$$\bullet \longrightarrow \bullet \longrightarrow \bullet \cdots \bullet \longrightarrow \bullet$$

of at least n worlds, so M cannot be finite.

It seems that very strong conditions are needed for **S4** miniatures. One way to enlarge the feasibility of **S4** miniatures is to restrict the set of consequences which has to be modelled (cf. [Le88]). For example, we can restrict the possible consequences to those of maximum modal depth d. As the heuristics from the beginning of section 8.3 indicates, given such a threshold there obviously are **S4** miniatures. This is, of course, a much more general point: regardless of the actual modal system, the construction will always yield a miniature for consequences upto depth d. And in fact, upto depth 1, there is no difference between the modal systems **S5**, $\mathbf{S5_{(m)}}$, **S4** or **T**, and the restricted miniatures will thus be equal for these systems.

Although *finite* characterization in general seems very difficult, we can study **S4** knowledge states from other points of view too. Concentrating on honest information we have to adapt the notion of *stability* for **S4**. [Ja91c] gives an abstract, yet elegant definition of *stability* for arbitrary normal systems, reminiscent of the work of Konolige and Moore.

Definition 8.9 (Jaspars) *A set T of formulas is* **S–stable** *iff*

1. *T contains the* **S**-*theorems* ($\vdash_{\mathbf{S}} \varphi \Rightarrow \varphi \in T$);
2. *T is deductively closed:* $\varphi \in T, \varphi \rightarrow \psi \in T \Rightarrow \psi \in T$;
3. $K[T] \cup \neg K[T^c]$ *is consistent with respect to* **S**.[26]

This definition is shown to be equivalent to the even more concise requirement that $T = K^{-1}[\Sigma]$ for some maximally **S**-consistent Σ. The new definition of stability is a correct generalization of Stalnaker's, since for **S5** both definitions amount to the same, and accord intuitions. The general notion can be applied to **S4**, singling out

[26] $\neg K[T^c] = \{\neg K\varphi \mid \varphi \notin T\}$

the **S4**-stable sets. Still we may wonder whether a definition of **S4**-stability can be found which is somewhat more transparant and closer to definition 8.6. What clearly has to change is clause *4* of the original definition, since this amounts to the feature of 'negative introspection', which is typical of **S5**. However, we want to maintain the correspondence between stable sets and theories validated by single models one way or the other, so we cannot simply drop condition *4*. What is implied by this condition for **S5** and has to hold for **S4** too is the property of what might be called *modal saturation*:

$$K\varphi \vee K\psi \in T \Rightarrow K\varphi \in T \text{ or } K\psi \in T$$

This leads to the following definition of **S4** stability:

Definition 8.10 *A set T of formulas is* **S4 stable** *iff*

1. *T contains the* **S4** *axioms;*
2. *$\varphi \in T, \varphi \rightarrow \psi \in T \Rightarrow \psi \in T$;*
3. *if $\varphi \in T$ then $K\varphi \in T$;*
4. *if $K\varphi \vee K\psi \in T$ then $\varphi \in T$ or $\psi \in T$;*
5. *T is consistent with respect to* **S4**.

The latter definition applies to all normal modal systems which extend **NK4**; in fact it is equivalent to definition 8.9 for all normal **NK4** extensions. Moreover, if we replace **S4** by **S5** definitions 8.10 and 8.6 are equivalent, so our definition correctly generalizes Stalnaker's.

Multi-modal logic and systems of various strength can also be fruitfully combined. One point in the network may deal with knowledge in a different way than others. To our knowledge, the heterogenous case has not been investigated yet. Still, this is what happens when people and computers exchange information.

8.7.3 The nature of the semantic structures

Here, in particular, partial models can be thought of. Although this multiplies the number of possible worlds and models in principle, it does not necessarily imply that the system becomes larger and slower: not every world has to be examined when all information is lacking with respect to a simple proposition; in fact partial miniatures will mostly be much smaller. This approach will be studied in the next chapter.

An alternative is to allow *infinite* miniatures on the formal level, and invent a smart way of finitely representing these infinities. Perhaps Vardi's knowledge structures may be interpreted in this way, i.e. as an implementation of possibly infinite Kripke models.

8.8 Conclusion

In this chapter we have investigated various ways of storing information in Kripke models. We focussed on attempts to push all information into one finite characterizing model. The obvious alternative to this idea is a procedure which supplies a counter-example for every non-consequence. This strategy will always succeed for the usual normal systems of modal logic, which possess the finite model property. So, this strategy is less exclusive. It is not even *a priori* clear whether this strategy is less efficient than mere model checking, but it is certainly less elegant: (essentially) the same things have to be done repeatedly. Therefore, the approach of devising 'miniatures', i.e. finite characterizing models, seems preferable.

The requirement of finiteness, modest though it may seem, poses considerable constraints on admissible sets of data in most cases of interest. The feasibility of what may be called *semantic* knowledge representation in fact turns out to depend on the background logic and the kind of models involved.

The main division here was between local models and global models. We showed that incomplete information cannot be characterized by a finite local model ('local miniature'), in the sense of verifying precisely the logical consequences of the data in a designated world. Under comparatively lenient conditions it appeared to be possible to characterize the data by a global model. In fact, various kinds of 'global miniatures' qualified. Each kind of miniature was shown to exist under certain necessary and sufficient conditions on deduction from the data.

The different kinds of miniatures can also be related directly. For example, a coarse miniature M, which characterizes the global consequences of the information, models α if and only if M is a global miniature for $K\alpha$. Local models can still be used when we do not want to describe *all* consequences, but only those of the form $K\varphi$. [27]

It should be noted that the condition of finiteness is immaterial for the existence of local miniatures: if there is a characterizing model, we may as well assume it to be finite, provided the logic has the finite model property. For global models finiteness really matters, since for introspective information the operation called 'disjoint union' in general produces a *countable* characterizing model. The point here is simply that the finite model property is a very common feature among standard modal systems, but logical finiteness, which appears to be required for global miniatures, is quite exceptional. For that reason it may be surprising that existence of local miniatures does not depend on logical finiteness: theorem 8.1 holds for **S4**, which is not logically finite.

Another formal difference between local and global models is that the former presuppose consistency of the data whereas the latter do not. Notice however that this difference is dominated by a definitional issue, viz. whether we are willing to allow empty (global) models or not. For the sake of simplicity we did allow empty models, but noticed that (global) consistency is a prerequisite for any concrete application.

Another question is whether a miniature is uniquely determined by the given

[27] This notion of *K-miniature* is related to our earlier notions by the fact that $K[D]$ has a non-empty tight miniature iff $K[D]$ has a K-miniature.

information. The answer to this is positive, in any case for **S5** and modulo reduction and isomorphy of models.

Existence and uniqueness of miniatures being guaranteed under the stated conditions, it has also been pointed out how miniatures can be effectively constructed from simple models, particularly for global and tight miniatures.

In this chapter we did not include complexity considerations.[28] Since the size of the models may be (super)-exponential in the number of atoms, the whole approach may seem intractable. However, recent research indicates that regularities and smart heuristics may make model checking quite efficient, even for very large structures, cf. [BC*90]. Still, reducing the size of the miniature is clearly advantageous. In the final chapter we will see whether partiality is capable of reducing the size of the miniatures.

[28]There are however some results concerning the size of the miniature in the next chapter.

Chapter 9

Partial models for knowledge-based systems

In the previous chapter we saw that knowledge representation by means of possible world models is feasible (at least for **S5**), but that the resulting models will be exceedingly large. We suggested that partial logic may help to reduce the size of the miniature. However, there is a paradox contained in this move. For although partial models may be smaller, there are more partial models than classical models, for a propositional atom may receive 3 truth-values instead of 2 in each world. Moreover, bringing to mind the general procedure leading to miniatures, in the weaker partial logic there will be *more* formulas that do not follow from the given information, hence more counterexamples are needed, including the classical ones, which jointly lead to a much *larger* miniature. Although this construction still works, we will find that usually there will exist a miniature that is much smaller than the 'canonical' one indicated above.

As we saw in chapters 6 and 7, partial epistemic logic may have the advantage that it is closer to actual human knowledge. Perhaps it is now also possible to capture these weaker types of knowledge in finite models, such as those triggered by partial modal systems resembling **S4**, which was virtually impossible for classical miniatures.

As we saw in chapter 3 there are two prevailing perspectives on validity in partial semantics: verification and (non-)falsification. We use both perspectives in this chapter, in particular, *relative* verification and falsification. Most of what follows will be directed to *global* miniatures, but we also reconsider *local* models, because the set of formulas true in some situation does not have to be maximally consistent, so, on the syntactic side, the problem of theoretical completeness (cf. fullness), which obstructed local miniatures for incomplete information in the classical case, is avoided by partiality. To prepare the ground for another alternative approach, viz. that of circumscriptive miniatures, we redefine *stability* for partial logics.

We provide complexity results for some types of information, both for partial and for classical miniatures. Minimization, which was a rather trivial operation on classical models (merely identifying propositionally equivalent worlds within the same component), now is a productive non-trivial operation: indeed this is what may make partial miniatures (much) smaller than classical ones.

9.1 Introduction

Why go partial in knowledge representation? The possibility of representing knowledge in finite Kripke models was studied in chapter 8. The main reasons for the present research stem from results reported in that chapter. Finite representation in classical possible worlds models (i.e. Kripke structures with a bivalent truth assignment) turned out to be possible under certain conditions. For example, some piece of **S5**-knowledge α can be characterized by a 'global miniature' M (i.e. M verifies precisely all the **S5** consequences of α in each world) iff α is *introspective* ($\alpha \vdash_{\mathbf{S5}} K\alpha$).

However, though the existence of classical miniatures encourages further research in this direction, some drawbacks of total models point at the need of partiality in model-theoretic knowledge representation. For, though the word 'miniature' indicates a tiny thing (reflecting our initial intention), the classical miniatures are by no means small.[1] To be more specific, we will calculate the size of one type of classical **S5** miniatures below.

Moreover, most positive results were obtained for **S5**, which is, in some sense, the simplest non-trivial modal logic. For the epistemic logics **S4** (which is somewhat closer to human knowledge) and $\mathbf{S5_{(m)}}$ (which is proper for the case of many agents reasoning according to **S5**) there seem to be no equally positive results. In particular, in chapter 8 it was proved that simple *incomplete* **S4**-knowledge cannot be modelled in finite classical Kripke models. A similarly negative result for $\mathbf{S5_{(m)}}$ showed that multi-agent knowledge cannot be modelled, without constraining the set of possible consequences to formulas of limited modal depth.

Finally, we noticed that modal systems such as **S4** do not account for the way in which human beings deal with knowledge — real agents are not perfect reasoners, therefore they will not know everything that follows from their knowledge. These observations lead to the following central questions:

- Can we improve upon the complexity of the representation of **S5**-knowledge? *A priori*, partial models seem proper to diminish the size of the miniatures.
- Can we represent the knowledge with respect to other epistemic logics, such as **S4** and $\mathbf{S5_{(m)}}$ by means of partial models?
- Can we represent the type of knowledge that is closer to the way in which human beings think, i.e. knowledge that accords to a weaker epistemic logic?[2]

[1] Perhaps this shift parallels the history of the word 'minim' (the next item in Longmans' International Reader's Dictionary): formerly a very short note of music, now quite a long one. Also, there is no consensus on the etymology of the word 'miniature', but the relation to Latin 'minim' is at least one of the possible sources.

[2] cf. [FH88] and chapters 6 and 7 here.

The answers to these points depend on the kind of evaluation.[3] Assuming a non-falsification perspective on valid consequence, we will show that the total miniatures are among the smaller ones. But under a verification perspective, the gain of partiality is more substantial since the miniatures will generally be much smaller and will follow the rules of somewhat weaker logics.

complexity of classical miniatures

Total miniatures soon become very large. In fact, the smaller the relative amount of information, the larger the model will be. Some of the worst cases are those of complete ignorance with respect to a number of propositional variables. Assume, for example, that the only information is Kp and the system is totally ignorant with respect to (only) three other atoms. Then the miniature will consist of 1024 worlds divided over 255 components. In general, if there is no information about r atoms, whereas the other atoms are completely known (i.e. either Kp_i or $K\neg p_i$ for, say, $i = 1, \ldots, n-r$), the number of worlds and components of the (smallest!) miniature is superexponential in r. More precisely,

Proposition 9.1 (size of miniature for simple ignorance)
A classical miniature modelling complete ignorance of r propositional variables and complete knowledge of the other variables, has $|C_r| = 2^{2^r} - 1$ components and a total number of $|W_r| = 2^{2^r+r-1}$ worlds.

Proof: First notice that the miniature is isomorphic to the model that characterizes zero information with respect to r atoms (simply drop the uniform specification for the known p_i out of the worlds). This model consists of all non-empty tight submodels[4] of the largest tight model for r atoms, which contains 2^r worlds. So there are $2^{2^r} - 1$ components. The total number $|W_r|$ of worlds in this model can be calculated by an easy combinatorial argument:[5]

$$|W_r| = \sum_{i=1}^{2^r} i \cdot \binom{2^r}{i} = \sum_{i=1}^{2^r} 2^r \cdot \binom{2^r - 1}{i-1} = 2^r \sum_{i=0}^{2^r-1} \binom{2^r - 1}{i} = 2^r \cdot 2^{2^r-1} = 2^{2^r+r-1}.$$

■

Notice that a special instance of this proposition is that in which the system has *no* knowledge at all of the n propositional variables. This 'worst case' analysis is even of some practical importance: a relational database can be relatively empty, that is, the number of atomic (predicate logical) formulas may be quite large, whereas the number of known facts small, and the miniature consequently gigantic (if $r \sim 100$, $|W_r| \sim 10^{10^{30}}$). As we will see, especially with such *simple ignorance*, partial miniatures have a dramatically better performance: given the right perspective, the model will consist of just one world!

[3]See the chapters 3 and 4 in this book.

[4]See chapter 8.

[5]Alternatively, Johan van Benthem has suggested that the number $|W_r|$ may be understood as follows: each of the 2^r state–descriptions occurs in 2^{2^r-1} components (equalling the number of subsets of other state descriptions to which the state is attached).

9.2 F–miniatures

First we will give the definition of an F–miniature, 'F' for non-falsification or falsifiability. Recall from chapter 3 that $M, s \not\dashv \varphi$ means that φ is not false in s according to model M, $M \not\dashv \varphi$ that $M, s \not\dashv \varphi$ for *every* s in M, and $D \not\dashv \varphi$ that $M, s \not\dashv \varphi$ for *every* model M and situation s such that $M, s \not\dashv \delta$ for *all* $\delta \in D$. So the slash in the consequence relation of relative falsifiability has a fixed meaning and does *not* indicate non-consequence.

Definition 9.1 *M is an* **F–miniature** *for D iff M is finite and for each φ: $M \not\dashv \varphi \Leftrightarrow D \not\dashv \varphi$.*

Or, equivalently, M is an F–miniature iff

- M is finite,
- $M \not\dashv D$,
- $M \not\dashv \varphi \Rightarrow D \not\dashv \varphi$.

Until further notice we will concentrate on models with an equivalence accessibility relation. Moreover, we assume the models to be coherent, i.e. the truth-value assignment can be defined or undefined, but not overdefined. Then the logic of the F–inferences is essentially **S5** without the (propositional and modal) *ex falso* rules (but with the rules of *tertium non datur*). This logic will be called **S5*** henceforth.[6]

We can give a syntactic criterion for the existence of F–miniatures. Both result and proof resemble the analogous case for classical miniatures.

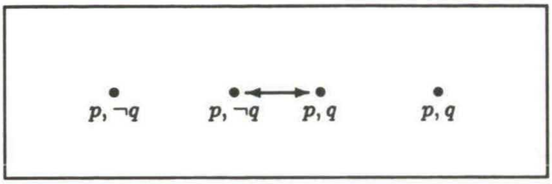
Theorem 9.1 *D has an F–miniature iff $D \vdash_{\mathbf{S5^*}} \bigwedge\!\!\!\bigwedge K[D]$.*

Proof: (Cf. the proof of theorem 8.2.) Notice that **S5*** induces a relation of derivational equivalence $\dashv\vdash$, which corresponds to F–equivalence (i.e. $\varphi \dashv\vdash \psi$ iff $M, s \equiv\!\!| \varphi \Leftrightarrow M, s \equiv\!\!| \psi$ for all M, s). With respect to this equivalence the logic is finite (corollary 4.2) and has the FMP (theorem 4.12). So the miniature is the finite disjoint union of finite counter-examples to non-consequences. Using the generation lemma, it is easy to show that this is an F–miniature for introspective D. ∎

This result demonstrates that for the usual data, such as epistemic formulas of the form $K\alpha$, model-theoretic representation is feasible. It does not display the form of the miniature, nor how to arrive at a *minimal* model. In fact, what does an F–miniature look like? In some cases an F–miniature may be a total model, as the following example will illustrate.

Example 9.1 (simple ignorance)
Let $Prop = \{p, q\}$ and $D = \{Kp\}$. The minimal total F–miniature for D is the model:

[6]See chapter 4. Recall that **S5*** is characterized by the set of rules $\mathbf{M^*} \cup \{\mathbf{T_r}, \mathbf{5_r}\}$, where $\mathbf{T_r}$ and $\mathbf{5_r}$ stand for $K\varphi \vdash \varphi$ and $\neg K\varphi \vdash K\neg K\varphi$ and their respective contrapositives.



So, the idea that invoking non-falsification always produces a reduction of the model (by dropping either positive or negative information) turns out to be wrong. We can, of course, *add* partialized components, or, more precisely, despecifications modulo finite equivalence.[7] In the present case this amounts to copying worlds within a component, reconnect them to the component and finally omit specifications. Partialization will lead to larger models, but, more importantly, the resulting miniatures will be equivalent.

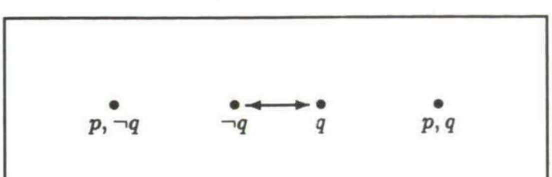
Proposition 9.2 *Adding partialized components to an F–miniature results in another F–miniature (for the same information).*

Proof: given some F–miniature M for D with component N, let $N'\dot{\sqsubseteq}N$, then $M \oplus N'$ is equivalent to M, for there are finite L, L' such that $N' \equiv L' \sqsubseteq L \equiv N$, and so (1) $M \oplus N'$ will be finite; (2) suppose for some $\delta \in D$ and s in M: $N', s \mathrel{\equiv\!\!|} \delta$, then by equivalence and persistence $N, s \mathrel{\equiv\!\!|} \delta$, thus $M, s \mathrel{\equiv\!\!|} \delta$, which contradicts $M \not\mathrel{\equiv\!\!|} D$. So we obtain $N \not\mathrel{\equiv\!\!|} D$, and therefore $M \oplus N' \not\mathrel{\equiv\!\!|} D$; (3) if $M \oplus N' \not\mathrel{\equiv\!\!|} \varphi$ then surely $M \not\mathrel{\equiv\!\!|} \varphi$, and so $D \not\mathrel{\equiv\!\!|} \varphi$. ■

The proposition licenses an optimalization procedure: an F–miniature can be minimalized by dropping components which are partializations of other components. So, an F–miniature thus minimalized will usually consist of more or less total components.

Now for those data that can be modelled by a total miniature, we may also consider whether we can supply a truly partial miniature by *replacing* a total component by its different partializations.

Example 9.2 (simple ignorance, continued)
Reconsider $D = \{Kp\}$ (cf. example 9.1). A minimal F–miniature for D might be obtained from the total miniature by omitting some of the literals. Notice however that contracting the middle component to the singleton p, will not do: then $Kq \vee K\neg q$ will be non-falsified, but $Kq \vee K\neg q$ is not a consequence of Kp. Likewise, dropping both occurrences of 'p' in the central component would produce the model

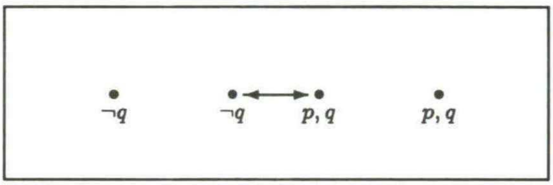


[7] So, $N_1\dot{\sqsubseteq}N_2$ iff for some finite N_3, N_4: $N_1 \equiv N_3$, $N_2 \equiv N_4$ and $N_3 \sqsubseteq N_4$. Here $\sqsubseteq$ expresses extension of valuation for the same frame. Persistence with respect to valuation extension was shown in chapter 4.

This model non-falsifies $K\neg p \vee Kq \vee K\neg q$, which is not an F–consequence of Kp. Similarly, dropping 'p' in the right-hand world of the middle component admits the non-consequence $\neg Kp \vee Kq \vee K\neg q$. In fact, as the reader may check (warning: this is tedious labour), none of the occurences of 'p', 'q' and '$\neg q$' may be omitted without loss of characterization.

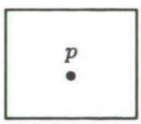
These examples suggest that minimal total F–miniatures are unique, up to isomorphism. The examples and the previous proposition may also suggest the generalization that classical (**S5**)- miniatures for some set of data D are always F–miniatures for D as well. Though tempting, the latter is not true.

Example 9.3 *Consider the data $D = \{Kp, K(p \to q)\}$. A total model that verifies D will also verify Kq. Consequently, essentially the only classical* **S5** *miniature for D will be the singleton model verifying both p and q. But Kq is not an F–consequence of D, roughly because* **S5*** *does not contain* Modus Ponens. *So the singleton model is not an F–miniature for D. A small F–miniature for D is:*



Another example of the incongruity of classical and V–miniatures may be more transparant. The point is that for inconsistent data, F–consequence and **S5**-consequence diverge widely.

Example 9.4 *For the data $\{Kp, Kq, K\neg q\}$ the minimal F–miniature is:*



As in the previous example, there is no total F–miniature for this set of data: a total model that non–falsifies q and $\neg q$ in each world should verify q and $\neg q$ in each world, thus has to be the empty model. But the empty model also non-falsifies $K\neg p$, which does not follow from the data in **S5***.

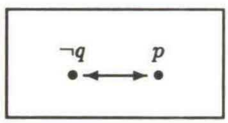
From these comparisons between (partial and total) F–miniatures and (total) classical miniatures some generalizations are induced:

- A total F–miniature for D is also a classical miniature for D.
- If D has a minimal F–miniature that is partial, it has no total F–miniature.

The first generalization is easily proved. Looking for a kind of converse notice that examples 9.3 and 9.4 show that classical miniatures may not be total F–miniatures, in fact that there is information that has a classical miniature, but no total F–miniature. So the following question emerges:

> *Is consistent information modelled by an F–miniature iff there is a classical miniature?*

From the left to the right this holds trivially (even without consistence): If D has an F–miniature then $D \vdash_{S5^*} \bigwedge K[D]$ and so $D \vdash_{S5} \bigwedge K[D]$, thus D has a classical miniature. For the other direction, notice we cannot leave out the consistency requirement; a trivial counter-example runs as follows: $p \wedge \neg p$ has the empty classical miniature, but cannot have an F–miniature, since $p \wedge \neg p \nvdash_{S5^*} K(p \wedge \neg p)$. This observation can be transformed into a genuine counter-example: Consider $\alpha = (p \wedge \neg p) \vee Kq$. α is consistent and classically equivalent to Kq, and therefore **S5** introspective. However α is not **S5*** introspective, for $K\alpha$ is falsified in the left-hand world of the following model, but α is not.



To conclude this section, we notice that in ordinary circumstances F–miniatures will not be of much practical importance. The reason for this is that though partiality may leave truth values open, leaving out many propositional specifications in model structures for incomplete knowledge would lead to wrong results under the present perspective on validity: simply too many formulas would become valid. We therefore turn to the other perspective on validity.

9.3 V–miniatures

The notion of 'V–miniature' is similar to 'F–miniature', with verification ('V') instead of non-falsification. The definitions of global verification ($M \models \varphi$), and relative verification (also called 'strong consequence') are obvious.[8]

Definition 9.2 *M is a* **V–miniature** *for D iff M is finite and for each φ: $M \models \varphi \Leftrightarrow D \models \varphi$.*

Or, equivalently, M is a V–miniature iff

- M is finite,
- $M \models D$,

[8] See chapters 3 and 4 for details.

- $M \models \varphi \Rightarrow D \models \varphi$.

As noticed in chapter 4, the set of V–consequences is usually considerably smaller than the set of normal consequences; so the V–miniature may have to represent less, which is enabled by the possibility of underspecifying worlds for their propositional contents.

In this section we still restrict ourselves to coherent models with an equivalence accessibility relation. For relative verification the inference rules will again be somewhat weaker than the usual **S5** ones. Call this logic, which is also a variant of good-old **S5**, now with the *ex falso* rules but without *tertium non datur*: **S5**$^+$, 'coherent verificational **S5**'.[9]

We may repeat the question of correspondence: which syntactic or deductive qualities of D enable its verificational representation by a finite partial model? Existence of V–miniatures is warranted by the already familiar syntactic condition of (deductive) introspection:

Theorem 9.2 *D has a V–miniature iff* $D \vdash_{S5^+} \bigwedge\!\!\!\bigwedge K[D]$.

Proof: (Cf. the proof of the previous theorem, now in a more symbolic fashion.) The condition is clearly necessary, but also sufficient. To show the latter, notice that **S5**$^+$ is logically finite and has the FMP. If Φ is the set of disjunctive normal forms of modal degree 1 (see section 4.6.3), then Φ is finite, and the disjoint union of counter-examples of non-consequences produces

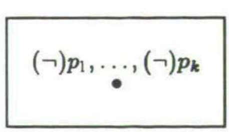
$$M = \bigoplus\{N \mid N \models D, N \text{ tight and reduced } \& \ N \not\models \varphi \text{ for some } \varphi \in \Phi\}.$$

It is easily checked that M is a V–miniature for D. ∎

Notice that this theorem guarantees existence of the V–miniature for introspective information, but does not produce a *minimal* V–miniature. As a matter of fact, the model produced in the proof will be usually (much) larger than the classical miniature. But in many cases smaller models can be obtained, as the examples below will demonstrate.

The gain of relative verification becomes clear in cases where only part of the propositional variables are known. Recall that for such *simple ignorance* the F–miniature amounts to a classical model of superexponential size (in the number of unknown variables).

Example 9.5 (simple ignorance)
Assume complete information about $p_1, \ldots, p_k$ (i.e. Kp_i or $K\neg p_i$ for each $i = 1, \ldots, k$), and complete ignorance of the rest. This set of data is modelled by the singleton miniature M: [10]

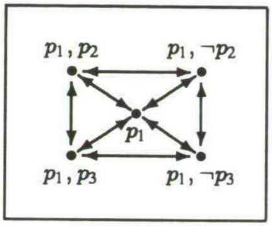


[9]See chapter 4. **S5**$^+$ is characterized by $\mathbf{M}^+ \cup \{\mathbf{T}_r, \mathbf{5}_r\}$.

[10]The correctness of this and the following miniatures is proved in the appendix.

This example is a case of 'simple ignorance': nothing is known about p_i for $i > k$, even $D \not\vdash \neg K p_{k+1} \vee \neg K \neg p_{k+1}$. It contrasts with types of 'strong ignorance' in which, for example, there is full knowledge about $p_1, \ldots, p_k$ but $D \vdash \neg K p_i \wedge \neg K \neg p_i$ for all $i > k$.

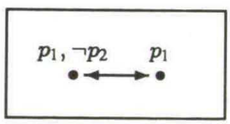
Example 9.6 (strong ignorance)
Let $D = \{Kp_1, \neg Kp_2, \neg K\neg p_2, \neg Kp_3, \neg K\neg p_3\}$. *D is represented by the following minimal model:*



Notice on the one hand that in cases of 'strong ignorance' the information can be strengthened further: in the above example e.g. $D \not\vdash \neg K(p_2 \vee p_3)$. Yet adding such a formula implies an increase of knowledge, so it seems fair to say that the ignorance would have decreased.

In between simple and strong ignorance there are intermediate cases of semi-strong (or, partial) ignorance. Here is a paradigm.

Example 9.7 (semi-strong ignorance)
Assume that $D = \{Kp_1, \neg Kp_2\}$. *D is minimally represented by:*



Now adding (negative) information such as $\neg Kp_i$ has a remarkable effect: it *reduces* the size of the classical miniature, but *magnifies* the size of the partial V–miniature somewhat. To wit, assume the initial information Kp_1. As we saw in the introductory section, the classical miniature for the atoms p_1, p_2, p_3 has 32 worlds (distributed over 15 components). Adding $\neg Kp_2$ leads to some reduction: for this case of semi-strong ignorance the classical miniature contains 28 worlds (12 components). The final addition of $\neg K\neg p_2$, $\neg Kp_3$ and $\neg K\neg p_3$ (strong ignorance) involves a classical miniature of 20 worlds (and 7 components). As the above examples show, the number of worlds of partial V–miniatures for these cases of simple, semi-strong and strong ignorance are 1, 2, and 4, respectively (and just one component in each case).

More generally, (semi-)strong ignorance can be captured by tight V–miniatures of polynomial, in fact even *linear* size, whereas their classical counterparts need a *superexponential* amount of worlds.

Proposition 9.3 (size of miniatures for (semi-)strong ignorance)
The minimal V–miniature modelling (semi-)strong ignorance with respect to r atoms and complete knowledge of the others has at most $2r + 1$ situations (in 1 component). A corresponding classical miniature requires at least 2^{2^r-1} worlds, divided over at least 2^{2^r-2} components (if $r > 0$).

Proof: the examples above of (semi-)strong ignorance obviate that the largest minimal V–miniature will be the one for strong ignorance of r atoms, where $2r + 1$ situations suffice: apart from the known literals, which occur in all situations, one containing p_i and one containing $\neg p_i$ for each unknown p_i and one which is empty.

For the classical case, notice that strong ignorance now gives the *smallest* miniature. Its largest component will have 2^r worlds (state–descriptions). Each set containing more than half of these state–descriptions will necessarily satisfy both p_i and $\neg p_i$ for all unknown p_i (for if not than there can be no more than 2^{r-1} states in the set), and thus will constitute a component of the miniature. So we have that

$$|C_r| \geq \sum_{2^{r-1}+1}^{2^r} \binom{2^r}{i} = \frac{1}{2} \cdot 2^{2^r} - \frac{1}{2} \binom{2^r}{2^{r-1}} \geq 2^{2^r-2}$$

if $r > 0$. A similar calculation for the number of worlds shows [11]

$$|W_r| \geq \sum_{2^{r-1}+1}^{2^r} i \cdot \binom{2^r}{i} = 2^r \sum_{2^{r-1}}^{2^r-1} \binom{2^r-1}{i} = 2^{2^r+r-2} \geq 2^{2^r-1}$$

if $r > 0$. (the last estimation also holds for the borderline case $r = 0$). ∎

This proposition illustrates our point that partial models are superior to classical models for at least two reasons: first, they are usually much smaller and, second, more natural since they tend to grow when information is added. Classical miniatures, on the other hand, are rather clumsy in describing knowledge. Total Kripke structures may be said to model ignorance rather than knowledge, which they are supposed to. One of the additional advantages of partial models is that, since simple ignorance does not need to be modelled, only 'relevant' propositional variables (which occur in the data) have to be taken into account. Consequently, we omit the specification of *Prop* in examples of V–miniatures.

Since (semi-)strong ignorance is a rather usual type of incomplete knowledge this is also of some practical importance. Moreover, the resulting miniatures are linearly dependent on the number of partly known propositional variables, consequently checking whether a formula is true can be executed in polynomial time, even with respect to $n = |Prop|$ and $|\varphi|$. Now polynomial time complexity is quite usual, also for classical models, then meaning: polynomial in $|\varphi|$ and the number of worlds $|W|$,[12] which, as we have seen may be of the order 2^{2^n}. Though [BC*90] shows that in some cases

[11] The summation may be replaced by a perhaps more insightful argument: more than half of the subsets of state–descriptions containing some particular state will contain at least half of the state–descriptions. This also yields the number $2^r \cdot \frac{1}{2} \cdot 2^{2^r-1}$.

[12] I am indebted to Edith Spaan for patiently explaining me the relevant complexity argument.

even large models can be processed, we conjecture that this will be impossible for, say, (semi-)strong ignorance.[13]

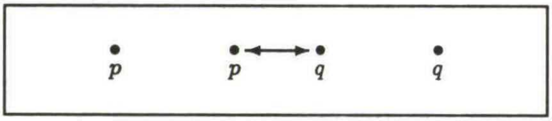
The above examples provided V–miniatures consisting of a single component. More complex types of incomplete knowledge may require several components, however. Before discussing some more involved examples, let us find the proper analogue for proposition 9.2. In order to reduce its size, we need to know how V–miniatures are related to partialization. Again the relation $\dot{\sqsubseteq}$ between components of a V–miniature involves a minimalization procedure, but now in the reverse direction. So, we can partly complete components by extending specification of worlds, after which equally specified worlds may be identified. Completization will also lead to larger models, but the resulting miniatures will be equivalent.

Proposition 9.4 *Adding partially completed components to a V–miniature results in another V–miniature (for the same information).*

Proof: similar to proposition 9.2. ■

By proposition 9.4 a V–miniature can be minimalized by dropping components which are essentially partial completions of other components. Therefore a minimal V–miniature will consist of 'most unspecified' components. This optimization is especially useful for more complex cases.

Example 9.8 (honest disjunctive knowledge)
The smallest V–miniature for $K(p \vee q)$ is:



In comparison with the classical miniature[14]the above model is still small: the classical **S5** *miniature is a graph consisting of 12 vertices and 6 edges, divided into 7 components.*

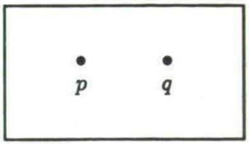
Although the last miniature consists of 3 components, the model as a whole represents a proper piece of information. One can succesfully declare to know *only* this or that; then the miniature has to be restricted to its central component. In so-called *dishonest* knowledge such a consistent circumscription is impossible: for example, one cannot consistently claim to know only whether this or that.[15]

[13]For example, if $r = 10$ then a partial miniature has at most 21 worlds, yet the classical miniature will have something of the order of $2^{2^{10}} \approx 10^{300}$, which exceeds the estimated number of protons and neutrons in the visible part of the universe ...

[14]Cf. example 8.3.

[15]See sections 8.5, 8.6, and the conclusion of chapter 8 for a more profound exposition and comparison of the two approaches, and section 9.5 here.

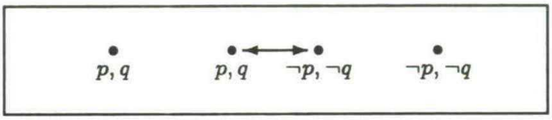
Example 9.9 (dishonest disjunctive knowledge)
The minimal V–miniature for $Kp \vee Kq$ happens to be the model:



Comparison with the classical miniature[16] *again points at a considerable reduction: the classical* **S5** *miniature has 7 worlds (and 2 edges, 5 components).*

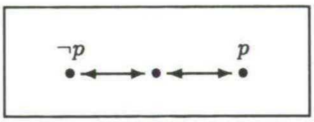
All the previous examples of incomplete information revealed V–miniatures which were (much) smaller than the classical ones. This suggest the generalization that representing incomplete knowledge by V–miniatures will always be more efficient. By inspection of a stronger kind of incomplete information it is shown that this generalization is not true.

Example 9.10 *The minimal V–miniature for $K(p \leftrightarrow q)$ is:*

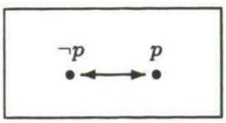


At first sight, things may even get worse in that V–miniatures may be larger than classical ones.

Example 9.11 *The minimal V–miniature for $\{\neg Kp, \neg K\neg p\}$ is:*



The classical miniature with respect to Prop $= \{p\}$ has one world less:



Before jumping to conclusions, note that proposition 9.3 indicates that larger V–miniatures are quite exceptional: for (semi-) strong ignorance the last example is the only case in which the linear growth function exceeds the superexponential one. Of course, other types of information may require larger miniatures. However, this does not give an increase of the theoretical complexity: after reduction, but even before minimalization triggered by persistence, the number of components is at most $2^{3^n} - 1$

[16] Cf. example 8.4

and the total number of worlds $3^n \cdot 2^{3^n-1}$ (cf. proposition 9.1). We have seen that minimalization usually cuts these numbers down considerably. Moreover, we believe that the actual absolute excess will be limited.

Given the extensive comparison between partial and classical miniatures in this section, we can evidently pose the same question as at the end of section 9.2: is existence of a V–miniature for (consistent) information equivalent to existence of a classical miniature. Actually, since $\mathbf{S5}^+$ contains the *ex falso* rules, the consistency requirement is immaterial now. Again the implication holds in one direction: if D has a V–miniature, D is $\mathbf{S5}^+$ introspective, and $\mathbf{S5}^+$ is contained in $\mathbf{S5}$, so D is $\mathbf{S5}$ introspective as well, and therefore has a classical miniature. In the other direction the implication is clearly false: $p \vee \neg p$ has a classical miniature (the same as for $\emptyset$), but no V–miniature, since $p \vee \neg p \nvdash_{\mathbf{S5}^+} K(p \vee \neg p)$.

9.4 Local miniatures

A perhaps more obvious semantic approach to characterize knowledge would involve Kripke's original *local* models. The idea is to have a designated world, from which the evaluation starts, containing the facts, and the accessible worlds containing the knowledge of these facts, and the knowledge of this knowledge, etcetera.

This approach was rejected in chapter 8 because only complete information can be described in this way. More precisely, only D which were *complete* and *consistent* theories qualified. We blamed bivalence in the root world for this obnoxious behaviour. Now, giving up bivalence, the hope of finding new possibilities for local miniatures is reviving. So, let us consider local F– and V– miniatures. To that purpose, replace M in definitions 9.1 and 9.2 by $\langle M, s\rangle$. We will discuss both kind of local miniatures separately.

Local F–miniatures

Like their global partners, the knowledge modelled by local F–miniatures does not have to be consistent. This oddity is illustrated by example 9.4, which also shows that if the global miniature is a singleton model, it coincides with the local miniature. Moreover, local F–miniatures do not capture honest disjunctive knowledge. For assume that, for example, $K(p\vee q)$ has a local miniature $\langle M, w\rangle$. Then $M, w \not\dashv K(p\vee q)$, so $M, w \not\dashv p$ or $M, w \not\dashv q$, but neither p nor q are consequences of $K(p \vee q)$. The conditions for local F–miniatures appear to be very strong.

Theorem 9.3
D has a local F–miniature iff D is both complete and saturated.[17]

Proof: analogous to theorem 8.1, using part of the canonical model by filtration over D + its subformulas. The suitability of saturation follows from the Henkin completeness proof of theorem 4.6 in section 4.4. ∎

[17]Completeness here expresses $D \vdash \varphi$ or $D \vdash \neg\varphi$, saturation $D \vdash \varphi \vee \psi \Rightarrow D \vdash \varphi$ or $D \vdash \psi$.

Local V–miniatures

Since (global) V–miniatures are more efficient then F–miniatures, we may hope for more success in the case of local V–miniatures than we experienced for local F–miniatures. However notice that disjunctive knowledge is still troublesome: $K(p \vee q)$ has no local V–miniature. The conditions for local V–miniatures are still very strong.

Theorem 9.4
D has a local V–miniature iff D is both consistent and saturated.

Proof: analogous to the previous theorem. ∎

9.5 Alternative: circumscription

Instead of *describing* knowledge as we did before, we may equally well be engaged in *circumscribing* knowledge: i.e. to model what one *only* knows. A classical model M circumscribes α when M is essentially the largest (tight) model for α. In this area the central syntactic notion connected to global models is *stability*. Following essentially [Ja91c], we take a stable set to be the theory of a global tight model. Of course the evaluation type matters, which is reflected here in the choice of the background logic ($\mathbf{S5}^+$ or $\mathbf{S5}^*$): [18]

Definition 9.3 *A set T of formulas is* $\mathbf{S5}^+$*–stable if* $T = \{\varphi \mid M \models \varphi\}$, *and* $\mathbf{S5}^*$*–stable if* $T = \{\varphi \mid M \not\dashv \varphi\}$ *where M is some non-empty tight model with an equivalence accessibility relation.*

An equivalent definition in syntactic terms is:

Proposition 9.5 *T is* $\mathbf{S5}^+$*–stable iff* $T = K^{-1}[\Sigma] = \{\varphi \mid K\varphi \in \Sigma\}$ *for some consistent saturated* $\mathbf{S5}^+$*–theory* Σ. *T is* $\mathbf{S5}^*$*–stable iff* $T = K^{-1}[\Sigma]$, *for some full saturated* $\mathbf{S5}^*$*–theory* Σ. [19]

Proof: ($\Rightarrow$) If T is $\mathbf{S5}^+$–stable then for some tight M: $T = \{\varphi \mid M \models \varphi\}$. Let s be a situation in M, and take $\Sigma = \{\varphi \mid M, s \models \varphi\}$. Then Σ is easily checked to be a consistent, saturated $\mathbf{S5}^+$-theory. Moreover, $T = K^{-1}[\Sigma]$ since $\varphi \in T \Leftrightarrow M \models \varphi \Leftrightarrow M, s \models K\varphi \Leftrightarrow K\varphi \in \Sigma$. ($\Leftarrow$) If $T = K^{-1}[\Sigma]$ for a consistent saturated $\mathbf{S5}^+$–theory Σ, then the Henkin completeness proof shows that Σ can be embedded in the canonical model $\mathcal{M}$ such that for the generated submodel $\mathcal{M}_\Sigma$ of $\mathcal{M}$: $\mathcal{M}_\Sigma, \Sigma \models \varphi \Leftrightarrow \varphi \in \Sigma$, which implies that T is $\mathbf{S5}^+$–stable. The proof for $\mathbf{S5}^*$–stability is similar. ∎

The novel definition may also be recast in more traditional terms.

[18] Definition 9.3 can easily be generalized for an *arbitrary* partial semantics.

[19] For saturated $\mathbf{S5}^*$–theories, fullness amounts to non-emptiness, cf. section 3.3.

Proposition 9.6 *T is* **S5**$^+$*–stable iff*
T is a consistent **S5**$^+$*–theory such that $K[T] \nvdash_{\mathbf{S5}^+} K[T^c]$ iff*

1. *T is an* **S5**$^+$*-theory (i.e. closed w.r.t. $\vdash_{\mathbf{S5}^+}$);*
2. *if $\varphi \in T$ then $K\varphi \in T$;*
3. *if $K\varphi \vee K\psi \in T$ then $\varphi \in T$ or $\psi \in T$;*
4. *$\varphi \notin T$ for some φ.*

Proposition 9.7 *T is* **S5***–*stable iff*
T is a non-empty **S5***–*theory such that $K[T] \nvdash_{\mathbf{S5}^*} K[T^c]$ iff*

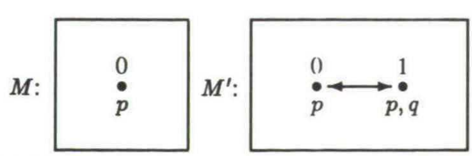
1. *T is an* **S5****-theory;*
2. *if $\varphi \in T$ then $K\varphi \in T$;*
3. *if $K\varphi \vee K\psi \in T$ then $\varphi \in T$ or $\psi \in T$;*
4. *if $\varphi \notin T$ then $\neg K\varphi \in T$.*

Proof: by means of the generalized Lindenbaum lemma (proposition 4.6). ∎

The clauses in the last equivalent of the propositions above bear a considerable resemblance to those for **S4**.[20] Notice that clause *2* expresses a kind of introspection, *3 modal saturation* and *4* non-emptiness, of T^c for **S5**$^+$ and of T for **S5***, modulo the other clauses.

Circumscription of knowledge involves a largest model (a last element in the ordering of supermodels). The corresponding stable set containing the information therefore has to be minimal, in some sense. In accordance with the usual terminology, let us call the data *honest* if such circumscription is possible. The double perspective of models and formulas gives rise to various notions of honesty, even for the verificational, c.q. **S5**$^+$ approach.

One subtlety of partial logic is that, different from the classical **S5** case, an **S5**$^+$ or **S5***–stable set is not fully determined by its subset of propositional formulas. For example, $\neg Kp \in T$ does not imply whether or not $p \in T$ for **S5*** and whether $\neg p \in T$ for **S5**$^+$. Here is a concrete counter-example for **S5**$^+$–stability. Consider $T = \text{TH}(M)$ and $T' = \text{TH}(M')$:



[20] Cf. [Ja91c] and chapter 8 here.

Since $\text{TH}(M',0) \subset \text{TH}(M',1)$, $\text{TH}(M') \cap \text{pL} = \text{TH}(M',0) \cap \text{TH}(M',1) \cap \text{pL} = \text{TH}(M',0) \cap \text{pL} = \text{TH}(M) \cap \text{pL}$. But clearly $\text{TH}(M) \neq \text{TH}(M')$ for $\neg K \neg q \in \text{TH}(M') - \text{TH}(M)$. There is a dual counter-example for **S5***–stability. This suggest that honesty may be defined as existence of a minimal stable extension, where minimality is just ordinary set inclusion. We will leave the elaboration of this point to another occasion.

9.6 Below and beyond partial S5

Within classical logic the paradigm cases of cautious extension or variation of the S5 system are **S4** and $\mathbf{S5_{(m)}}$. The possibilities for miniaturization of information turned to be extremely limited for these background logics. For the epistemic logic S4 (advocated by Hintikka) complete information can be represented in a singleton model. In chapter 8 we conjectured that incomplete information does not have **S4** miniatures; this claim has been verified for simple and semi-strong ignorance. The situation for $\mathbf{S5_{(m)}}$ is even worse: consistent information (whether 'complete' or not) does not have $\mathbf{S5_{(m)}}$ miniatures. Is the situation for partial logic similarly distressing? To study this in some detail, we will focus on V–miniatures in the rest of this section.

$\mathbf{S4^+}$ miniatures

Partiality slightly improves the chances for $\mathbf{S4^+}$ miniatures: both complete knowledge and simple ignorance can be modelled by singleton V–miniatures. A simple induction proof shows that $\mathbf{S4^+}$ and $\mathbf{S5^+}$ have the same inferences from this kind of information. Consequently, the miniature of example 9.5 still qualifies with respect to $\mathbf{S4^+}$.

For (semi–)strong information we find the opposite situation. For example, contrasting to example 9.7, $\{Kp, \neg Kq\}$ has no $\mathbf{S4^+}$ miniature, for suppose M would qualify. Then $M \models \neg Kq$, so $M \models K\neg Kq$. However, $Kp, \neg Kq \not\models K\neg Kq$.

More generally, *negative* information appears to obstruct possible miniatures. More challenging are the cases of *positive* partial knowledge, such as $K(p \vee q)$. We believe that a characterizing model requires chains of unlimited length, but this has not been proven with formal rigour yet. On the other hand, the 'dishonest' formula $Kp \vee Kq$ appears to have the same miniature as in example 9.9.

$\mathbf{S5^+_{(m)}}$ miniatures

Here the situation is worse. One of the points is that for a multi-agent logic there can be no complete information *without* contingent common knowledge, which can not be expressed in the simple modal language. And without an operator for common knowledge, we cannot obtain miniatures: the models will always be too strong. In fact the proof of the nonexistence of $\mathbf{S5_{(m)}}$ miniatures for consistent data can simply be transposed to the realm of partial semantics. On the other hand, adding an operator C for common knowledge may not solve all problems. Though it is clear that C–

introspection is a necessary condition for the existence of $\mathbf{S5}^+_{(\mathbf{m})}$ miniatures, it is not obvious that the condition is sufficient.

9.7 Conclusion

In this chapter we considered the feasibility of partial miniatures. It is argued that the notion of validity that pays in this respect is relative verification, rather than falsification. In the latter case the gain in cases of simple incomplete information is nil, compared to classical miniatures. In the former case the gain is quite good: for some types of incomplete information, the resulting miniatures are indeed very small. The explanation for this difference was connected to the different direction of what might be called 'miniature persistence': F–miniatures are closed under partialization of the model, V–miniatures under (partial) completion. Consequently, the smallest F–miniatures are the most specified (and therefore usually big), the smallest V–miniatures the least specified (thus small). Therefore, V–miniatures are often quite efficient.

As in the classical case we saw that existence of partial miniatures corresponds to the information α being introspective, in the sense that $\alpha \vdash K\alpha$, where of course $\vdash$ is indexed for the appropriate partial modal system. Yet, where the disjoint union construction essentially supplied the classical miniature, for partial miniatures minimization is a productive and non-trivial next step, especially for V–miniatures.

Different from what might be expected, the notion of local miniature is still unsuitable for characterizing, for example, disjunctive knowledge. The other alternative of partial circumscription is related to the definitions of stability with respect to the partial systems $\mathbf{S5}^+$ and $\mathbf{S5}^*$. Perhaps it is possible to give smaller partial models, when circumscribing instead of describing knowledge.

We also had a quick look at some other modal systems, such as the partial counterparts of **S4** and $\mathbf{S5}_{(\mathbf{m})}$. In the former case, some progress has been observed: unlike for classical semantics, simple ignorance can be modelled now. Yet, for other types of information in the **S4**-like case, and in the $\mathbf{S5}_{(\mathbf{m})}$-like case in general, there does not seem to be much improvement. This calls for a restriction on the modal depth of formulas that are checked on the model. Then again partiality improves the size of the miniatures.

Although there are, of course, in general differences when we compare the **S5** consequences to the $\mathbf{S5}^+$ consequences, the obtained minimization is a satisfying product of the collaboration of modal logic and partial semantics.

Appendix: correctness of V–miniatures

correctness of example 9.5 (simple ignorance)

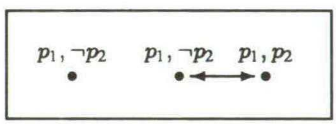
Proof: Clearly $M \models D$, and assume that $M \models \varphi$. Let $N, s \models D$ for some model N and situation s, then for the generated submodel N_s also $N_s \models D$. Thus $M \dot{\sqsubseteq} N_s$, for M can be obtained from N_s by omitting all p_i specifications for $i > k$ and then identify equally specified situations. Therefore $N_s \models \varphi$, so $N, s \models \varphi$, and consequently $D \models \varphi$. ■

correctness of example 9.6 (strong ignorance)

Proof: Correctness and minimality of this miniature can be shown by constructing the semi-lattice of verifying tight models, partially ordered by $\dot{\sqsubseteq}$, which has the displayed miniature as its bottom element. This, however, is a laborious exercise. An easy argument may do just as well: Notice that every tight model verifying D will consist of situations which at least contain p_1, whereas some situations should contain p_2, $\neg p_2$, p_3, and $\neg p_3$. Now the given model can be strengthened to any verifying model, and thus minimally characterizes D (cf. proposition 9.4 for formal justification). ■

correctness of example 9.7 (semi-strong ignorance)

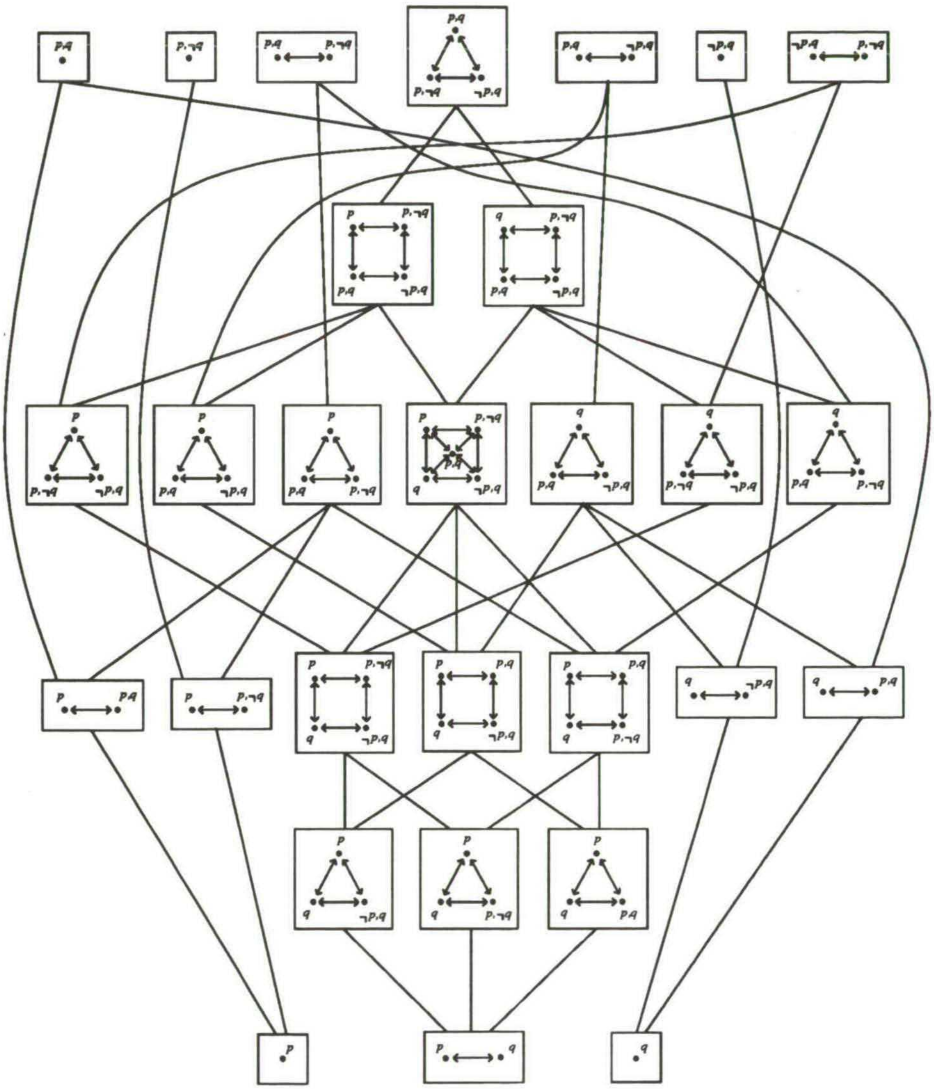
Proof: Here a construction of the semi-lattice of verifying components is feasible. In fact there are two ways to generate this structure. One may inspect the 8 minimal models which have p_1 in each world and possibly also p_2 or $\neg p_2$, and check whether $\neg K p_2$ holds in them. Then the obtained models are ordered for $\dot{\sqsubseteq}$, and the displayed miniature appears to be the bottom of 3 element semi-lattice. But one may also start with the *classical* miniature for D: its components will verify D in the partial sense too.



Then by weakening (i.e. partializing and possibly duplifying and identifying situations) the minimal partial miniature is obtained in the end. The latter method often turns out to be more efficient. ■

correctness of example 9.8 (honest disjunctive knowledge)

Proof: it suffices to draw the graph of all (reduced) tight V–models for $K(p \vee q)$, ordered with respect to $\dot{\sqsubseteq}$ (indicated by lines), where the uphill components are more specific and thus may be omitted. (see next page) ■


p,q
p,¬q
p,q p,¬q
p,q p,¬q ¬p,q
p,q ¬p,q
¬p,q
¬p,q p,¬q
p p,¬q p,q ¬p,q
q p,¬q p,q ¬p,q
p p,¬q ¬p,q
p p,q ¬p,q
p p,q p,¬q
p p,¬q p,q q ¬p,q
q p,q ¬p,q
q p,¬q ¬p,q
q p,q p,¬q
p p,q
p p,¬q
p p,¬q q ¬p,q
p p,q q ¬p,q
p p,q q p,¬q
q ¬p,q
q p,q
p q ¬p,q
p q p,¬q
p q p,q
p
p q
q

Samenvatting

Zowel[1] voor de representatie van kennis in menselijke individuen, als voor de representatie van die kennis in computersystemen kunnen er grote voordelen verbonden zijn aan het gebruik van partiële logica. In partiële logica is het niet zo dat elke bewering waar of onwaar is: de waarheidswaarde kan ook open gelaten worden. Hoewel dit de semantiek van de logica op zich ingewikkelder maakt, kan kennisrepresentatie door middel van partiële logica adequater en efficiënter geschieden dan voorheen met klassieke logica. Het doel van dit onderzoek is te toetsen in hoeverre partialiteit hier inderdaad uitkomst biedt.[2]

Na een algemene inleiding vangt het proefschrift aan met een technisch deel waarin verschillende systemen van partiële logica worden onderzocht, zowel op uitdrukkingskracht als op geldige redeneerpatronen. Gezien de beoogde toepassingen wordt hier de taal van de modale propositielogica beschouwd, waarin naast connectieven ('niet', 'en', 'of', ...) ook modale operatoren ('het is noodzakelijk/mogelijk dat') optreden, die later de rol van kennisoperatoren gaan spelen ('A weet dat', 'A gelooft dat', ...).

Omdat de combinatie van partiële en modale logica nog grotendeels onontgonnen gebied was, is eerst de partiële propositielogica (dus zonder modale operatoren) bestudeerd. Met betrekking tot uitdrukkingskracht is hier onderzocht welke semantische condities met welke propositielogische talen corresponderen. Ook ten aanzien van geldige redeneerpatronen zijn de mogelijkheden systematisch onderzocht. Een complicerende factor hierbij is het aantal vrijheidsgraden dat de partiële logica kenmerkt: het aantal waarheidswaarden, het soort geldigheid, het soort gevolgtrekkingsrelatie, en de interpretatie van de connectieven. Ons beperkend tot een standaardinterpretatie van connectieven, hebben wij van de andere aspecten de belangrijkste mogelijkheden in kaart gebracht. Hierbij bleken een aantal al bekende logica's, zoals de klassieke propositielogica met partiële semantiek, de logica van sterk gevolg en de basale relevantielogica, vrijwel vanzelf hun plaats binnen deze systematiek te krijgen.

Vervolgens wordt de uitbreiding naar de modale taal bekeken. De complexere strukturen die nodig zijn om de modale taal adequaat te interpreteren laten helaas meer toe dan in de modale taal uitdrukbaar is. Om die reden wordt de uitdrukkingskracht van de modale taal niet gerelateerd aan de modellen zelf, maar aan het soort predikatenlogica dat ook met deze modellen geïnterpreteerd zou kunnen worden. Gelukkig bleek hier een equivalentierelatie tussen de modellen, bekend als 'bisimulatie', voldoende om het typerende predikaatlogische fragment te karakteriseren.

De systematiek ten aanzien van geldigheid, zoals ontwikkeld voor het propositionele geval, bleek geheel en al van toepassing op de modale taal. De belangrijkste noties van geldigheid zijn beschreven in deductieve systemen (vergelijk de axiomastelsels uit de wiskunde), die, voor zover ons bekend, nieuw zijn. De zg. volledigheidsbewijzen

[1] Since the contents of this summary is predictable from the foregoing text, it may serve to some readers as a first introduction to a fragment of the Dutch language. The introduction and conclusions of the chapters together form a summary in English.

[2] In deze samenvatting beperken we ons tot de hoofdlijn van het onderzoek, voorbijgaand aan veel op zich belangrijke technische details en zijsporen.

van deze beschrijvingen vormen de technische kern van dit proefschrift.

Kenmerkend voor het meest gangbare soort modellen van modale logica is de aanwezigheid van een toegankelijkheidsrelatie tussen werelden. Deze toegankelijkheidsrelaties vormen een extra vrijheidsgraad voor de modale semantiek, en door het stellen van beperkingen daarop is het mogelijk algemene verschillen tussen bij voorbeeld weten en geloven te verdisconteren. Met het oog op latere toepassingen worden een aantal uit deze beperkingen resulterende modale systemen nader onderzocht. Ook hier blijken klassieke resultaten zoals volledigheidsstellingen, de eindige-modeleigenschap, normaalvormen en (in enkele gevallen) logische eindigheid opgang te doen.

In het tweede deel van dit proefschrift wordt de partiële logica toegepast op het gebied van menselijke taal en kennis. Hierbij blijken partiële systemen daadwerkelijk grote voordelen te bieden ten aanzien van het modelleren van expliciete menselijke kennis en bewustzijn, als ook bij het beschrijven van taaluitingen en het verklaren van daarbij optredende verschijnselen.

Bij het laatste wordt een regel geformuleerd die de bedoeling heeft de 'toegevoegde waarde' van het *uiten* van een bewering juist in te schatten. Als belangrijkste proefgeval geldt hierbij de paradox van Moore: waarom is de *zin* 'Het regent maar ik geloof dat niet', hoewel op zich logisch consistent, toch merkwaardig om te uiten? De regel van de toegevoegde pragmatische waarde verklaart deze en vergelijkbare gevallen door de juiste epistemische modaliteit aan de uiting toe te kennen en te laten zien dat aldus in het geval van Moore's paradox alsnog een logische contradictie ontstaat. Hoewel we deze afleidingen in klassieke logica geven, wordt ook aangetoond dat zulks ook mogelijk is in een veel zwakkere partiële logica.

Bij het aangrenzende onderwerp van mogelijke logica's voor meer realistische vormen van weten en geloven wordt een expliciete relatie gelegd met bewustzijn: we zullen in de regel niet alle gevolgen van onze kennis kennen, omdat we ons van die gevolgen niet bewust hoeven zijn. Klassieke logica's verplichten daarentegen wel tot kennis van alle gevolgen; dit staat bekend als het probleem van *logische alwetendheid.*

Eerst worden enige totale logica's beschouwd die aanpassingen zijn van de klassieke logica maar nog steeds een totale (niet-partiële) semantiek kennen. Een van de voorstellen uit de literatuur, nl. de logica van 'algemeen bewustzijn', wordt door een eenvoudige ingreep zo gegeneraliseerd dat deze in feite alle vormen van bewustzijn kan uitdrukken, zowel (intuïtief gesproken) 'logische' als 'onlogische' vormen. Aangezien dit resultaat zeer wel als een Pyrrhus-overwinning gezien kan worden (het kader is zo algemeen geworden dat het in feite leeg is), worden ook andere logica's in ogenschouw genomen, zowel totale als partiële.

Vooral de laatste zijn wat dit betreft van belang, omdat het partiële kader als het ware vanzelf, zonder de kunstgrepen uit de totale semantiek, al te sterke eigenschappen voor weten of geloven kan vermijden. Dit lukt evenwel alleen als de notie van geldig gevolg wordt uitgelegd als sterk gevolg (*relatieve verificatie*): als in een coherente situatie de premissen waar zijn, dan moet de conclusie ook waar zijn. Dan ontstaat echter het probleem dat deze partiële semantiek niet meer de klassieke tautologieën oplevert, hetgeen onwenselijk wordt geacht: u gelooft bij voorbeeld dat het regent, of u gelooft dat niet, en de semantiek moet deze geldigheid verantwoorden. Om aan deze complicatie het hoofd te bieden wordt een combinatie van klassieke en partiële semantiek voorgesteld. De resterende vormen van logische alwetendheid die de partiële benadering onverlet laat, kunnen, zo men dit verkiest, worden geëlimineerd door de al eerder beschouwde zeef van bewuste formules bovenop het gewone interpretatiemechanisme te plaatsen.

Tenslotte wordt in het derde deel van dit proefschrift de geschiktheid van partiële modellen voor het representeren van kennis in computers nader onderzocht. Voor bepaalde soorten (introspectieve) kennis is de representatie in eindige modellen mogelijk, maar tenminste voor onvolledige kennis leidt dit bij gebruik van een klassieke interpretatie tot bijzonder grote (superexponentiële) modellen.

De partiële semantiek met relatieve verificatie kan hier uitkomst bieden: ontbrekende kennis kan dan door de afwezigheid van specificaties worden verantwoord. Toch schuilt hier een addertje onder het gras: in principe overtreft zowel het aantal als de grootte van de partiële modellen dat van de klassieke modellen. De oplossing van deze paradox is dat de partiële modellen weliswaar groter kunnen zijn, maar dat we op die grootte sterk kunnen bezuinigen. Slechts in uitzonderlijke grensgevallen zal het partiële model (iets) groter zijn het klassieke. In de meest voorkomende gevallen kan de representatie nu echter aanzienlijk efficiënter; voor eenvoudige vormen van onvolledige kennis zelfs met modellen waarvan de grootte lineair afhangt van het aantal feiten waar slechts gedeeltelijke informatie over bestaat.

Bibliography

[Åq64] Åqvist, L. - 'A solution to Moore's paradox', *Philosophical Studies XV*, pp. 1–5, 1964

[Ac68] Aczel, P. - 'Saturated intuitionistic theories', in: H. Schmidt, K. Schütte & H. Thiele (eds.) *Contributions to Mathematical Logic*, North-Holland, Amsterdam, 1968

[Ba81] Barwise, J. - 'Scenes and other situations', *The Journal of Philosophy, 78*, no. 7, 1981

[BP81] Barwise, J. & J. Perry - 'Situations and attitudes', *The Journal of Philosophy, 78*, no. 7, 1981

[Be77] Belnap, N. - 'A useful four–valued logic', in: Dunn & Epstein (eds.) *Modern Uses of Multiple–valued Logic*, pp. 8–37, Reidel Dordrecht, 1977

[vB84a] van Benthem, J. - 'Partiality and nonmonotonicity in classical logic', *Logique et Analyse* 29, pp. 225–247, 1986

[vB84b] van Benthem, J. - 'Correspondence theory', in: Gabbay & Günthner (eds.) *Handbook of Philosophical Logic*, volume 2, pp.167-247, 1984

[vB84c] van Benthem, J. - 'Questions about quantifiers', *Journal of Symbolic Logic* 49, pp. 443-466, 1984

[vB85] van Benthem, J. - *Modal Logic and Classical Logic*, Bibliopolis, Napoli, 1985

[vB86] van Benthem, J. - *Essays in Logical Semantics*, Reidel, Dordrecht, Holland, 1986

[vB88] van Benthem, J. - *A Manual of Intensional Logic*, CSLI Lecture Notes No. 1, 2nd edition, Stanford CA, 1988

[vB90] van Benthem, J. - 'Modal logic as a theory of information', in: J.Copeland (ed.) *Proceedings of the Prior Memorial Colloquium, Christchurch 1989*, to appear with Oxford University Press, Oxford UK

[Be56] Beth, E.W. - 'Semantic construction of intuitionistic logic', in: *Mededelingen der* KNAW, Afdeling Letterkunde, deel 19, no. 11, pp. 357–388, 1956

[Bl52] Black, M. - 'Saying and disbelieving', *Analysis 13*, pp.25–33, 1952; reprinted in: M. Black, *Problems of Analysis*, Cornell University Press, Ithaca, 1954 and Greenwood Press, Westport, 1971

[Bl86] Blamey, S. - 'Partial logic', in: Gabbay and Günthner (eds.) *Handbook of Philosophical Logic*, volume 3, Reidel, Dordrecht, 1986

[BD84] Božić, M. & K. Došen - 'Models for normal intuitionistic modal logics', *Studia Logica*, pp. 217–245, 1984

[Bo37] Bochvar, D.A. - 'Ob odnom tréhznacnom iscislenii i égo priménéniik analizu paradoksov klassicéskogo rassirénnago funkcional'nogo iscisléniá', *Matématiceskii sbornik 4* (46), pp. 287–308, 1937; English translation 'On a three–valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus' in: *History and Philosophy of Logic* 2, pp. 87–112, 1981

[BK79] Bunt, H. & A. van Katwijk - 'Dialogue acts as elements of a language game', in: Zonneveld & Weerman (eds.) *Linguistics in the Netherlands 1977–1979*, pp. 264–282, van Gorcum, Assen, 1979

[Bu90] Bunt, H. - *Modular incremental modelling of belief and intention*, ITK Research Report 20, Katholieke Universiteit Brabant, Tilburg, 1990

[BC*90] Burch, J., E. Clarke, D. Dill, J. Hwang & K. McMillan - 'Symbolic model checking: 10^{20} states and beyond', *Proceedings of the 5^{th}* IEEE *Symposium on Logic in Computer Science*, pp. 428–439, 1990

[BS84] Bull, R. & K. Segerberg - 'Basic modal logic', in: Gabbay and Günthner (eds.), *Handbook of Philosophical Logic*, volume 2, 1984

[Ca64] Castañeda, H.-N. - 'Review of [Hi62]', *Journal of Symbolic Logic*, pp. 132–134, 1964

[Ch80] Chellas, B. - *Modal Logic. An Introduction.* Cambridge University Press, Cambridge (UK), 1980

[Cu63] Curry, H.B. - *Foundations of Mathematical Logic*, McGraw–Hill, New York, 1963

[Du86] Dunn, J. - 'Relevance logic and entailment', in: Gabbay and Günthner (eds.) *Handbook of Philosophical Logic*, volume 3, Reidel, Dordrecht, 1986

[vE91] van Eijck, J. (ed.) - *Logics in AI*, Proceedings JELIA'90, (les Journées Européennes sur la Logique en Intelligence Artificielle), Lecture Notes in Artificial Intelligence 478, Springer-Verlag, Berlin, 1991

[FH88] Fagin, R. & J. Halpern - 'Belief, awareness and limited reasoning', *Artificial Intelligence 34*, pp. 39–76, 1988

[FHV90] Fagin, R., J. Halpern & M. Vardi - 'What is an inference rule?', *Proceedings Jerusalem Conference on Information Technology* pp. 391–401, 1990. Expanded version as IBM Research Report RJ 7865, 1990

[FHV91] Fagin, R., J. Halpern & M. Vardi - 'A model-theoretic analysis of knowledge', *Journal of the* ACM 38:2, pp. 382–428, 1991

[FH*87] Fenstad, Halvorsen, Langholm & van Benthem - *Situations, Language and Logic*, Reidel, Dordrecht, 1987

[Fi75a] Fine, K. - 'Vagueness, truth and logic', *Synthese* 30, pp. 265–300, 1975

[Fi75b] Fine, K. - 'Normal forms in modal logic', *Notre Dame Journal of Formal Logic* 16, pp. 229–237, 1975

[vF66] van Fraassen, B. - 'Singular terms, truth-value gaps, and free logic', *Journal of Philosophy* 63, pp. 481–495, 1966

[Fu71] Furberg, M. - *Saying and Meaning*, Basil Blackwell, Oxford, 1971

[Ga79] Gazdar, G. - *Pragmatics. Implicature, Presupposition, and Logical Form*, Academic Press, New York , 1979

[GG90] Gillet, E. & P. Gochet - *La logique de la connaissance. Le problème de l'omniscience logique*, Université de Liège, Séminaire de Logique et d'Epistémologie, 1990

[Gi74] Gilmore, P.C. - 'The consistency of partial set theory without extensionality', *Proceedings of Symposia in Pure Mathematics*, volume 13, part II, American Mathematical Society, Providence, 1974

[Gö33] Gödel, K. - 'Eine Interpretation des intuitionistischen Aussagenkalküls, *Ergebnisse eines mathematischen Kolloquiums* 4, 1933

[Go87] Goldblatt, R. - *Logics of Time and Computation*, CSLI Lecture Notes No. 7, Stanford CA, 1987

[GMR85] Goldwasser, S., S. Micali & C. Rackoff - 'The knowledge complexity of interactive proof systems', *Proceedings of the 17^{th}* ACM *Symposium on Theory of Computation*, pp. 291–304, 1985

[GP89] Goranko, V. & S. Passy - *Using the universal modality: gains and questions*, unpublished manuscript, 1989

[Gr68] Grice, P. - 'Utterer's meaning, sentence-meaning, and word-meaning', *Foundations of Language 4*, 1968, pp. 1–18; reprinted in: J. Searle (ed.) *The Philosophy of Language*, Oxford University Press, Oxford (UK), 1971

[Gr75] Grice, P. - 'Logic and conversation', in: Cole & Morgan (eds.) *Syntax and Semantics 3: Speech Acts*, Academic Press, New York, 1975

[Gr78] Grice, P. - 'Further notes on logic and conversation', in: P. Cole (ed.) *Syntax and Semantics 9: Pragmatics*, Academic Press, New York, 1978

[HM85] Halpern, J. & Y. Moses - 'Towards a theory of knowledge and ignorance', in: Kr. Apt (ed.), *Logics and Models of Concurrent Systems*, Springer-Verlag, Berlin, 1985

[HMT88] Halpern, J., Y. Moses & M. Tuttle - 'A knowledge-based analysis of zero knowledge', *Proceedings of the 20th ACM Symposium on Theory of Computation*, pp. 132–147, 1988

[HV91] Halpern, J., & M. Vardi - 'Model checking vs. theorem proving: a manifesto', in: J. Allen (e.a. eds.) *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference*, 1991. Also appeared as IBM Research Report RJ 7963, 1991

[Ha77] Harnish, R. - 'Logical form and implicature', in: Bever, Katz and Langendoen (eds.) *An Integrated Theory of Linguistic Ability*, The Harvester Press, Hassocks, 1977

[Hi62] Hintikka, J. - *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca, 1962

[Hi75] Hintikka, J. - 'Impossible possible worlds vindicated', *Journal of Philosophical Logic* 4, pp.475-484 , 1975

[vdH91] van der Hoek, W. - *Systems for knowledge and belief*, in: [vE91, pp.267–281]

[vdHM88] van der Hoek, W. & J.-J. Meyer - *Possible logics for belief*, report IR-170, Vrije Universiteit, Amsterdam, 1988

[HK91] Huang, Z. & K. Kwast - 'Awareness, negation and logical omniscience', in: [vE91, pp.282–300]

[HC68] Hughes, G. & M. Cresswell - *An Introduction to Modal Logic*, Methuen, London, 1968

[HC84] Hughes, G. & M. Cresswell - *A Companion to Modal Logic*, Methuen, London, 1984

[Hu81] Humberstone, I. - 'From worlds to possibilities', *Journal of Philosophical Logic* 10, pp. 313–339, 1981

[Ja91a] Jaspars, J. - 'Theoretical circumscription in partial modal logic', in: [vE91, pp.301–316]

[Ja91b] Jaspars, J. - 'Fused modal logic and inconsistent belief', *Proceedings of* WOCFAI'*91*, Angkor, Paris, 1991 (also to appear in *Studia Logica*)

[Ja91c] Jaspars, J. - 'A generalization of stability and its application to circumscription of positive introspective knowledge', *Proceedings of the Ninth Workshop on Computer Science Logic*, Springer-Verlag, Berlin, 1991

[Ja92] Jaspars, J. - 'Normal forms in partial modal logic', *Proceedings of the XXXVIII Banach Centre Semester on Algebraic Methods in Logic and their Computer Science Application*, C. Rauszer (ed.), Warsaw, to appear

[KT89] Kammler, M. & E. Thijsse - *Something you might want to know about 'wanting to now'*, ITK Research Report 4, Katholieke Universiteit Brabant, Tilburg, 1989

[Ka83] Kamp, H. - *A scenic tour through the land of naked infinitives*, 139 pp., unpublished manuscript, 1983

[KM85] Keenan, E. & L. Moss - 'Generalized quantifiers and the expressive power of natural language', in: J. van Benthem en A. ter Meulen (eds.) *Generalized Quantifiers in Natural Language*, GRASS 4, Foris, Dordrecht, 1985

[Kl52] Kleene, S. - *Introduction to Metamathematics*, Van Nostrand, Princeton, 1952

[Ko86] Konolige, K. - 'What awareness isn't: a sentential view of implicit and explicit belief', *Proceedings of* TARK1 (Monterey CA), pp. 241–250, Morgan Kaufmann, 1986

[KL88] Kraus, S. & D. Lehmann - 'Knowledge, belief and time', *Theoretical Computer Science 58*, pp. 155–174, 1988

[Kr63] Kripke, S. - 'Semantical analysis of modal logic I', *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* 9, 1963

[Kr65a] Kripke, S. - 'Semantical analysis of modal logic II. Non-normal modal propositional calculi', in: Addison, Henkin & Tarski (eds.) *The theory of models*, pp. 206–220, North-Holland, Amsterdam, 1965

[Kr65b] Kripke, S. - 'Semantical analysis of intuitionistic logic I', in: *Formal Systems and Recursive Functions*, eds. J. Crossley & M. Dummett, North-Holland, Amsterdam, 1965

[Kr75] Kripke, S. - 'Outline of a theory of truth', in: *Journal of Philosophy* 72, pp. 690–716, 1975

[La87] Lakemeyer, G. - 'Tractable meta-reasoning in propositional logics of belief', *Proceedings of* IJCAI87 *(Milan)*, Morgan Kaufmann, Los Altos CA, pp. 402–408, 1987

[LL88] Lakemeyer, G. & H. Levesque - 'A tractable knowledge representation service with full introspection', in: M. Vardi (ed.) *Proceedings of* TARK2 (Monterey CA), pp. 145–159, Morgan Kaufmann, 1988

[La88] Langholm, T. - *Partiality, Truth and Persistence*, CSLI Lecture Notes No. 15, Stanford CA, 1988

[Le65] Lemmon, E. - 'Review of [Hi62]', *The Philosophical Review 74*, pp. 381–385, 1965

[Le78] Lenzen, W. - 'Recent work in epistemic logic', *Acta Philosophica Fennica* 30, pp. 1–219, 1978

[Le80] Lenzen, W. - *Glauben, Wissen und Wahrscheinlichkeit*, Springer-Verlag, Vienna/New York, 1980

[Le84a] Levesque, H. - 'A logic of implicit and explicit belief', *Proceedings of the National Conference on Artificial Intelligence*; expanded version as FLAIR *Technical Report 32*, Fairchild Laboratory of AI Research, Palo Alto CA, 1984

[Le84b] Levesque, H. - 'The logic of incomplete knowledge bases', *On Conceptual Modelling*, Brodie (e.a. eds.), Springer-Verlag, New York; reprinted in J. Mylopoulos & L. Brodie (eds.), *Readings in Artificial Intelligence and Databases*, Morgan Kaufmann, San Mateo (CA), 1989

[Le88] Levesque, H. - 'Logic and the complexity of reasoning', *Journal of Philosophical Logic* 17, pp. 355 – 389, 1988

[Le83] Levinson, S. - *Pragmatics*, Cambridge University Press, Cambridge UK, 1983

[Łu20] Łukasiewicz, J. - 'O logice trójwartościowej', *Ruch Filozoficzny 5*, pp. 170–171, 1920; English translation 'On three–valued logic' in: McCall - *Polish Logic*, Clarendon Press, Oxford, 1967

[Ly59] Lyndon, R. - 'Properties preserved under homomorphisms', *Pacific Journal of Mathematics 9*, pp. 143–154, 1959

[Ma54] Martin, N. - 'The Sheffer functions of 3–valued logic', *the Journal of Symbolic Logic*, pp. 45–51, 1954

[MD80] McDermott, X. & Y. Doyle - 'Non-monotonic logic I', *Artificial Intelligence 13*, pp. 41–72, 1980

[Mo68] Montague, R. - 'Pragmatics', in: R. Klibansky - *Contemporary Philosophy: a Survey*, La Nuovo Italia Editrice, Florence, 1968; reprinted in: Montague, *Formal Philosophy*, Yale University Press, 1974

[Mo12] Moore, G.E. - *Ethics*, London, 1912; reprinted by Oxford University Press, Oxford (UK), 1958

[Mo85] Moore, R. - 'Semantical considerations on non-monotonic logic', *Artificial Intelligence* 25, pp. 75–94, 1985

[Mo89] Morikawa, O. - 'Some modal logics based on a three-valued logic', *Notre Dame Journal of Formal Logic* 30, pp. 130–137, 1989

[Mo88] Moses, Y. - 'Resource-bounded knowledge', in: M. Vardi (ed.) *Proceedings of* TARK2 (Monterey CA), pp. 261–275, Morgan Kaufmann, 1988

[Mu89] Muskens, R. - *Meaning and Partiality*, thesis, Universiteit van Amsterdam, 1989

[Pa85] Patel–Schneider, P. - 'A decidable first-order logic of knowledge representation', *Proceedings* IJCAI85, 1985

[PW89] Pearce, D. & H. Wansing - 'On the methodology of possible worlds semantics, II: non-normal worlds and propositional attitudes', Gruppe für Logik, Wissenstheorie und Information an der Freie Universität Berlin, Bericht 2/89, 1989 (to appear in *Acta Philosophica Fennica*)

[Pe90] Penrose, R. - *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*, Oxford University Press, Oxford (UK), 1990

[Po21] Post, E. - 'Introduction to a general theory of elementary propositions', *American Journal of Mathematics 43*, pp. 163–185, 1921

[Ra82a] Rantala, V. - 'Impossible worlds semantics and logical omniscience', *Acta Philosophica Fennica 35*, pp. 106–115, 1982

[Ra82b] Rantala, V. - 'Quantified modal logic: non-normal worlds and propositional attitudes', *Studia Logica 41*, pp. 41–65, 1982

[Re69] Rescher, N. - *Many–valued Logics*, McGraw–Hill, New York, 1969

[RB80] Rescher, N. & R. Brandom - *The Logic of Inconsistency*, Basil Blackwell, London, 1980

[dR90] de Rijke, M. - 'The modal logic of inequality', ITLI, Universiteit van Amsterdam, 1990 (to appear in *Journal of Symbolic Logic*)

[Ro77] Rosenberg, I. - 'Completeness properties of multiple–valued logic algebras', D. Rine (ed.) *Computer Science and Multiple–valued Logic*, pp. 144–186, North–Holland, Amsterdam etc., 1977

[Ro85] Rosenschein, S. - 'Formal theories of knowledge in AI and robotics', *New Generation Computing* 3, pp. 345–357, 1985

[Sa91] Sánchez Valencia, V. - *Studies on Natural Logic and Categorial Grammar*, thesis, Universiteit van Amsterdam, 1991

[SJ*78] Schotch, P., J. Jensen, P. Larsen & E. MacLellan - 'A note on three-valued modal logic', *Notre Dame Journal of Formal Logic* 19, pp. 63–68, 1978

[Sc70] Scott, D. - 'Advice on modal logic', in K. Lambert (ed.) *Philosophical Problems in Logic*, pp. 143–173, Reidel, 1970

[Se67] Segerberg, K. - 'Some modal logics based on a three–valued logic', *Theoria* 33, pp. 53–71, 1967

[Se70] Segerberg, K. - 'Modal logics with linear alternative relations', *Theoria* 36, pp. 301–322, 1970

[Sł39] Słupecki, J. - 'Kryterium pełności wielowartościowych systemów logiki zdań' (A criterion of fullness of many–valued systems of propositional logic), *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, Classe III, volume 32, pp. 102–109, 1939

[St87] Stirling, C. - 'Modal Logics for Communicating Systems', *Theoretical Computer Science 49*, 1987

[Ta92] Tan, Y.H. - *Non-Monotonic Reasoning: Logical Architecture and Philosophical Applications*, thesis, Vrije Universiteit, Amsterdam, 1992

[Th85] Thijsse, E. - 'Counting quantifiers', in: J. van Benthem en A. ter Meulen (eds.) *Generalized Quantifiers in Natural Language*, GRASS 4, Foris, Dordrecht, 1985

[Th90a] Thijsse, E. - *Partial logic and modal logic: a Systematic Survey*, ITK Research Report 11, Katholieke Universiteit Brabant, Tilburg, 1990

[Th90b] Thijsse, E. - 'Partial propositional and modal logic: the overall theory', M. Stokhof & L. Torenvliet (eds.) *Proceedings of the 7th Amsterdam Colloquium*, volume 2, pp. 555–579, ITLI, Amsterdam, 1990

[Th90c] Thijsse, E. - *Definability in partial logic: the propositional part*, ITK *Research Report 24*, Katholieke Universiteit Brabant, Tilburg, 1990

[Th91a] Thijsse, E. - 'On total awareness logics (with special attention to monotonicity constraints and flexibility)', in: M. de Rijke (ed.) *Colloquium on Modal Logic 1991*, to appear as *Diamonds and Defaults, Studies in Logic, Language and Information*, volume 1, Kluwer, 1992 (to appear)

[Th91b] Thijsse, E. - *Simple models of knowledge, Proceedings* CLIN'91, Centrum voor Wiskunde en Informatica, Amsterdam, to appear

[Th92a] Thijsse, E. - 'Finite Kripke models for knowledge representation', in: R. van de Riet & R. Meersman (eds.), *Proceedings of the 1991 Workshop on Linguistic Instruments in Knowledge Engineering (*LIKE91*)*, Elsevier Science Publishers, Amsterdam, 1992

[Th92b] Thijsse, E., *Partial semantics for awareness*, ITK, Tilburg University, forthcoming Research Report, 1992

[Th68] Thomason, R.H. - 'On the strong semantical completeness of the intuitionistic predicate calculus', *The Journal of Symbolic Logic* 33:1, 1968

[Ur86] Urquhart, A. - 'Many valued logic', in: Gabbay and Günthner (eds.) *Handbook of Philosophical Logic*, volume 3, Reidel, Dordrecht, 1986

[Va86] Vardi, M. - 'On epistemic logic and logical omniscience', *Proceedings of* TARK1 (Monterey CA), pp. 293–305, Morgan Kaufmann, 1986

[Ve81] Veltman, F. - 'Data semantics', in: Groenendijk, Janssen & Stokhof (eds.) *Formal Methods in the Study of Language*, volume 2, Mathematical Centre Tracts 136, Amsterdam, 1981

[Ve85] Veltman, F. - *Logics for Conditionals*, thesis, Universiteit van Amsterdam, 1985

[Ve87] Veltman, F. - *Partiële Logica's* ('Partial Logics'), lecture notes, Universiteit van Amsterdam, 1987

[Vi84] Visser, A. - 'Four valued semantics and the liar', *Journal of Philosophical Logic 13*, pp. 181–212, 1984

[Wa89] Wansing, H. - 'Bemerkungen zur Semantik nicht-normaler möglicher Welten', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik 35*, pp. 551–557, 1989

[Wa90] Wansing, H. - 'A general possible worlds framework for reasoning about knowledge and belief', *Studia Logica 49*, pp. 523–539, 1990

[We84] Westerståhl, D. - 'Some results on quantifiers', *Notre Dame Journal of Formal Logic* 25, pp. 152–170, 1984

[Zw83] Zwarts, F. - 'Determiners: a relational perspective', in: A. ter Meulen (ed.) *Studies in Modeltheoretic Semantics*, GRASS 1, Foris, Dordrecht, 1983

EBURON DELFT

ISBN 90-5166-267-X